AD-A250 294

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

...ted to average 1 hour per response, including the time for reviewing instructions, searching existing data sources ...eviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1992, April	3. REPORT TYPE AND DATES COVERED Final Report June 91 - Oct 91

4. TITLE AND SUBTITLE Toward a Fuzzy Theory of Performance Measurement	5. FUNDING NUMBERS DAAL03-86-D-0001 62785A 790
6. AUTHOR(S) Mahan, Robert P.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Georgia Department of Psychology Athens, GA 30602	8. PERFORMING ORGANIZATION REPORT NUMBER --
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute Fort Knox Field Unit ATTN: PERI-IK Fort Knox, KY 40121-5620	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARI Research Note 92-28

DTIC ELECTE MAY 1 8 1992 S C D

11. SUPPLEMENTARY NOTES

Contracting Officer's Representative, Donald F. Haggard
Task was performed under a Scientific Services Agreement issued by Battelle, Research (Continued)

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE --

13. ABSTRACT (Maximum 200 words)

This report offers an alternative method for evaluating military unit performance. Fuzzy set theory is presented as a formal model of language expressions used in value judgments made by military experts. Further, it is suggested that fuzzy set theory may be used to connect the automated or instrumented physical measures taken from unit exercises with subjective judgments of expert military observers who interpret events in the framework of accepted concepts and principles of effective ground operations. Finally, the report documents a case study of a military domain of constructs wherein a sample of commanders identify three dimensions of communication and the semantic networks used to judge reports on these dimensions. The networks provide a sample group of interrelated terms that can be used to investigate the validity of fuzzy set operations for representing how expert observers describe and quantify aspects of communication performance.

14. SUBJECT TERMS Fuzzy sets Expert judgment Rating scales	Performance measurement Performance evaluation Military communications	15. NUMBER OF PAGES 61 16. PRICE CODE --

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std 239-18

ARI Research Note 92-28

11. SUPPLEMENTARY NOTES (Continued)

Triangle Park Office, 200 Park Drive, P.O. Box 12297, Research Triangle Park, NC 27709 and U.S. Army Research Office, P.O. Box 12211, Research Triangle Park, NC 27709.

ARI Research Note 92-28

Toward a Fuzzy Theory of Performance Measurement

Robert P. Mahan

University of Georgia

for

Contracting Officer's Representative
Donald F. Haggard

Field Unit at Fort Knox, Kentucky
Barbara A. Black, Chief

Training Systems Research Division
Jack H. Hiller, Director

April 1992

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

MICHAEL D. SHALER
COL, AR
Commanding

Research accomplished under contract
for the Department of the Army

University of Georgia

Technical review by

Billy L. Burnside

NOTICES

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992, April	Final Report June 91 - Oct 91

4. TITLE AND SUBTITLE
Toward a Fuzzy Theory of Performance Measurement

5. FUNDING NUMBERS
DAAL03-86-D-0001
62785A
790

6. AUTHOR(S)
Mahan, Robert P.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Georgia
Department of Psychology
Athens, GA 30602

8. PERFORMING ORGANIZATION REPORT NUMBER
--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute
Fort Knox Field Unit
ATTN: PERI-IK
Fort Knox, KY 40121-5620

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 92-28

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Donald F. Haggard
Task was performed under a Scientific Services Agreement issued by Battelle, Research
(Continued)

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT (*Maximum 200 words*)

This report offers an alternative method for evaluating military unit performance. Fuzzy set theory is presented as a formal model of language expressions used in value judgments made by military experts. Further, it is suggested that fuzzy set theory may be used to connect the automated or instrumented physical measures taken from unit exercises with subjective judgments of expert military observers who interpret events in the framework of accepted concepts and principles of effective ground operations. Finally, the report documents a case study of a military domain of constructs wherein a sample of commanders identify three dimensions of communication and the semantic networks used to judge reports on these dimensions. The networks provide a sample group of interrelated terms that can be used to investigate the validity of fuzzy set operations for representing how expert observers describe and quantify aspects of communication performance.

14. SUBJECT TERMS
Fuzzy sets — Performance measurement
Expert judgment — Performance evaluation
Rating scales — Military communications

15. NUMBER OF PAGES
61

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500 — Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18 298-102

11. SUPPLEMENTARY NOTES (Continued)

Triangle Park Office, 200 Park Drive, P.O. Box 12297, Research Triangle Park, NC 27709 and U.S. Army Research Office, P.O. Box 12211, Research Triangle Park, NC 27709.

ACKNOWLEDGMENTS

The author is indebted to David W. Bessemer for the time, patience, and expertise he provided throughout this project. It could not have been completed without his help. The author would also like to thank the commanders of the 194th Armored Brigade, Task Force 1-10 Cavalry, for their valuable participation and willingness to help with the completion of this project.

TOWARD A FUZZY THEORY OF PERFORMANCE MEASUREMENT

EXECUTIVE SUMMARY

Requirement:

This research examines the development of a formal system of expert military judgment that would lead to rules for operating on subjective linguistic-based assessments of training performance. The primary purpose of the research is to introduce the concept of a multiphase effort designed to develop a measurement theory for performance characteristics derived from exercising subject matter expertise. The report focuses on developing a measurement theory that will augment and extend automated performance measurement systems under development for device-based training assessment.

Procedure:

The first phase of this multiphase project was to characterize the shortfalls of current measurement techniques by demonstrating their tendency to obscure the meaning of expert military judgment. The argument is made that, without a formal method of classifying and operating using the natural language expressions that form the basis of many expert judgments of tactical performance, the true meaning of subject matter expertise will never be fully captured in the performance measurement process. To begin developing such a theory of measurement, a group of commanding officers was asked to help define the natural language syntax used when evaluating communications reporting performance by tank unit platoon leaders.

Findings:

The commanders generated a list of linguistic terms that afforded a reasonable degree of flexibility in grading the communication performance of platoon leaders. Findings related to commanders' assessment processes appear to indicate that many tactical activities require the imprecision of linguistic-based performance evaluation because of difficulties in precisely documenting the many dimensions of complex performance.

Utilization of Findings:

The results of this preliminary project lay some of the logical groundwork for developing a measurement system more compatible with the cognitive process of exercising expert military judgment. A measurement theory of the kind discussed in this report would offer a more sophisticated and valid method for modeling subjective military judgment and would increase the breadth and precision of device-based combined arms tactical training assessment procedures. When fully developed, the measurement methods discussed should have wide applicability to training innovations and be of interest to Army agencies responsible for testing and evaluating the effectiveness of training devices and simulators.

TOWARD A FUZZY THEORY OF PERFORMANCE MEASUREMENT

CONTENTS

LIST OF TABLES

LIST OF FIGURES

TOWARD A FUZZY THEORY OF PERFORMANCE MEASUREMENT

Introduction

Consider the conversation between two military experts describing a series of tactical events that they have just observed on a simulated battlefield. Expert 1 turns to Expert 2 and makes the point that mission effectiveness suffered because few reports were transmitted to command informing them of enemy contact. Expert 2 responds that he agrees and further indicates that several opportunities existed during the battle for transmitting tactical information. Here, Expert 1 clearly understood what he was saying to Expert 2. Similarly, Expert 2 understood what Expert 1 meant and actually extended the logic of Expert 1 by noting how many occasions existed for sending reports. In other words, this would seem a perfectly routine discussion between two trained military observers until one stopped to consider what the vague terms "few" and "several" mean. If, for example, the experts had been asked directly just how many reports constitute a "few" or how many occasions compose "several," they probably would have hesitated, then responded with "two" or "three" for "few," and "five" or "six," or perhaps "seven," for "several." Furthermore each expert would probably have generated approximately the same values. Although the experts may have differed some on the particular numbers given to specify "few" reports and "several" occasions, both experts probably would have viewed any of these alternative numbers as reasonably acceptable definitions.

This example illustrates the transmission of some vague, quantitative information between two individuals describing a potentially complex military event. One might argue that a great deal of quantitative information associated with military operations is vague in nature. A movement-to-contact operation may go "very well"; a sector may be defended "in depth"; an intelligence report may be "not very old"; an enemy force in contact may be "quite large." Not only do military personnel understand such statements; they also are able to manipulate and otherwise operate on these vague concepts.

There has been much interest over the years in the field of linguistics in documenting the vagueness in language and determining how one goes about quantifying meaning in natural language terminology. Although this area has been extensively studied (e.g, Lakoff, 1973), the focus of this interest changed when special mathematical operations became available for studying the vagueness in natural language concepts.

Fuzzy set theory defines concepts and techniques that provide a logic system to deal with logical relations that are too imprecise for classical mathematical techniques (Zadeh, 1973). Fuzzy set theory is an extension of classic set theory that relaxes the strong condition that an event be either in or out of a set, but not both. Fuzzy sets permit events to be partially included inside and outside a set simultaneously. The power seen in fuzzy set theory is that the concept of partial membership appears more compatible with human cognition than discrete choice, which conforms to the classic set theory approach to measurement (Schmucker, 1984; Smithson, 1987; Zadeh, 1973). A major feature of fuzzy set theory is that any system that can be quantitatively specified can contain both numeric and vague (linguistic) variables. Fuzzy operators on linguistic variables can be used similarly to nonfuzzy operators on numeric variables.

Nonfuzzy measurement systems typically rely on the axioms embodied in classic set theory and require that objects or events be uniquely categorized into well-defined sets. Further, they require that objects (things) or events and their properties must be classified as either belonging or not belonging to a given set of measurements, but not both. When a researcher imposes the notion that measurements can be uniquely assigned to sets in this manner (i.e., either belonging to the set A or not A), the researcher assumes that the individuals producing the measurements can make this distinction as well, and often in an intuitive way. For example, having individuals indicate the subjective level of some attribute of an object or event as a point on a rating scale is but one example of a classic measurement technique. Generating data in a manner that supports the axioms of classic analysis is assumed to correspond with the way in which the ratings were produced by the individuals under study. However, in many cases, the measuring device tends to extract data more exact than the subjective responses representing the corresponding measured human experiences (Polkinghorne, 1984).

A particular feature of nonfuzzy systems is the imposition of assumptions regarding the notion of uncertainty. Taking a decision-making viewpoint, the classic nonfuzzy approach in defining event uncertainty is that although specific sets of outcomes exist for a given action or set of actions, these outcomes may be unknown. However, implicit in the assumption of uncertainty is that there exists a random process that underlies the connection between actions and outcomes. Under this interpretation, a decision maker generates assessments regarding the membership or nonmembership of an event in some class or set of events. Here, uncertainty lies in not knowing to which set the event under consideration by the decision maker belongs.

However, the notion of fuzziness is distinctly different. Fuzziness is a function of not being able to precisely delineate among the groups of possible outcomes. Here, the decision maker is not able to precisely partition the state of the world into well-defined units. This appears to be more consistent with natural decision-making environments, where complexity is related to not knowing what the optimal courses of action are. As a simple example, consider the situation of assigning new cars to the set of "expensive cars." In the classic sense, uncertainty would be defined as not knowing to which set a new car might belong. However, after you examined the sticker price or asked the saleperson about the cost of the car, uncertainty would be elliminated. Either the car would meet the defining criterion and belong to the set of "expensive cars", or it would belong to the alternative set "not expensive cars". In contrast, consider the conceptual meaning of the term "expensive". At what dollar figure does a car abruptly transition from "expensive" to "not expensive". The argument here is that no exact dollar figure can be used to define a precise point of transition. Instead there is a boundry region that defines a gradual transition from expensive to not expensive. The decision maker will never be able to precisely determine whether a car is expensive or not, even after the salesperson indicates its cost.

Fuzzy set theory provides a possible solution to the methodological problems associated with assumptions regarding subjects' abilities to precisely document events. It takes into account the reality of the imprecision in human thought by allowing ranges of scores to be measured and translated into a single linguistic estimate. It is conceivable that fuzzy variables will be able to be used in statistical analyses in traditional ways, although more research is needed to verify this claim. While future work will likely require creating fuzzy statistical techniques that can be used to support fuzzy measurement, using more traditional statistics along with fuzzy measures means that current psychometric standards of validity and reliability can be applied to evaluate the potential of the fuzzy measurement process.

A particular application in which it is worth examining the usefulness of fuzzy sets is the issue of individual differences found among military experts in judgments derived from exercising their subject matter expertise to examine military systems and operations. One issue associated with expert judgment is typically viewed as the extent to which differences in military judgments are a function of genuine individual differences rather than artificial differences constrained or induced by the measurement procedures themselves.

Many of the methods used in military science for measuring performance by means of expert knowledge restrict an individual's responses both in terms of the content under study and the process by which it is measured. For example, measurement dimensions are typically defined and specified prior to any data collection efforts. This fact potentially limits the expert judge to measurement dimensions that appeal to the idiosyncratic biases of the experimenter. Further, typical experimental situations constrain the responses of an expert judge to a single choice along some prespecified measurement continuum. Guilford (1975), as well as others, have indicated that the constraints imposed on subjects by conventional measurement techniques may affect assessment of individual differences. Further, experimental evidence appears to confirm the notion that people learn to process and manipulate precise quantitative information in a "more-or-less" fashion, (Brehmer, 1973, 1976; Klienmuntz, 1985; Simon, 1978). This fact is the principle guiding current developmental efforts in analog display technology, which seeks to exploit the natural tendency of people to process quantitative information in an imprecise, approximating manner (Wickens, 1984). This imprecision of cognitive processing, in part, results from the fact that conceptual boundaries tend to be blurred across people even though fundamental conceptual meanings remain relatively constant (Neisser, 1967).

Fuzzy set theory adds an additional set of techniques that can be used to document complex systems that are composed of both numeric and linguistic information. It may provide a possible means whereby one can quantify the judgments of military experts expressed in their analyses and assessments of complex tactical operations. Specifically, this approach may hold promise for characterizing the complex performances found in simulator training environments. In this context, it would be very useful in being able to measure the meaning of words and phrases that make up expert military judgments of simulated battles, along with describing the reasoning process behind these judgments.

Objective

The objective of this report is to introduce the first phase of a multi-phase project to connect the theory of fuzzy sets with performance assessment and evaluation procedures currently used by the U.S. Army. The report will discuss some of the conceptual issues that surround assessing performance in complex military settings. Specifically, the discussion will focus on the use of military experts in interpreting tactical behavior of individuals and their units. Furthermore, the report highlights that the performance assessment process made by these experts contain both numeric and linguistic information. The report thus builds on the idea of applying the procedures of fuzzy sets to model the meaning of concepts and relations used by military experts in assessing military performance. The goal is to lay some of the ground work for establishing mechanisms that support integrating subjective and objective performance measures within the common framework of military theory. The research findings will ultimately be used to augment methods used by the Army to assess training effectiveness for device-based training systems.

The need for the Army to continue to pursue research and development of advanced measurement technologies will likely become greater in the future. This need is based primarily on the expanding role of high technology, device-based training programs. These programs are giving the Army the potential to create highly sophisticated, relatively inexpensive, simulated battlefield environments that can be used to train soldiers. With these new environments come new possibilities for measuring the effectiveness of simulator training by developing measures that relate to the task standards embedded in training doctrine.

Performance measurement systems and procedures have always played a pivotal role in training doctrines in the Army. A measure of performance is typically observed to be a term, quantity, or group of quantities, which are believed to summarize the behavior of soldiers and their units. Decision makers often use performance measures in order to: (a) provide training feedback to soldiers, (b) evaluate training needs, and (c) to manage various training systems.

The performance measures themselves are always intended, at the very least, to communicate information which will allow for rank ordering the various attributes and dimensions that compose military operations. Presumably, those who use the measure of performance can ignore the technical issues associated with how the measures were generated. The decision maker will instead make evaluations based on how the measures are rank ordered.

Decision makers do not generally consider the formal scaling properties of the measures they use. Instead the measures often become embedded in a kind of conversational vocabulary which frequently finds its way into both technical and nontechnical discussions. Some common examples relating to ground forces are shown in Table 1.

Table 1

Examples of Performance Measures

Time to Plan Mission	Range of Target Engagements
Time to Execute March	Distance Travelled During March
Accuracy of SPOT Report	Accuracy of Contact Report
Time to Plan FRAGOS	Number of FRAGOS Executed
Time to Execute Mission	Rate of March During Mission

Situational context often complicates the meaning assigned to particular performance measures. Nevertheless, there are certain common features of performance measures that allow decision makers to agree upon their appropriate use in particular situations. For example, evaluating the performance of a tactical road march would not typically include measures specifically useful in evaluating target acquisition and engagement, although both sets of measures may be based on a similar metric, such as units of time. Therefore, performance measures are almost always constrained by the context in which they are used. In this sense, observable events must undergo higher order transformations in order to add, among other things, contextual meaning to the measures.

The context dependent transformations made on performance measures typically produce performance indices which combine both quantitative and linguistic information. In many situations, value judgments, which are primarily linguistic-based, are mixed with numerical data. Further, it is the value judgment portion of the measure that forms the linkages between the numerical data, the tactical context, and the military constructs necessary to infer meaning to a given military event. Value judgments often include linguistic qualifiers such as "Good" timing, "Costly" maneuver, "Informative" SPOT report, etc.

However, traditional Army policy in establishing guidelines for performance measurement systems is largely based on a discrete classification system (e.g., qualified/unqualified, go/no go, untrained/needs training/trained). These methods are thus crude in the sense that they do not offer a means for dealing with ambiguity, vagueness, bias, or degrees of opinion that usually characterize the interpretational complexity of military operational environments. The discrete classification methodology, for example, contrasts with how subject matter experts (SMEs) perform in practice when their duties are based on detailed descriptions and analyses of critical incidents within the context of certain military constructs and doctrine (Hiller, 1987). In this sense, the range and quality in responses necessary to support expert judgments of observable tactical events is often artificially constrained to discrete categories. This discrete classification forces the expert to make very precise distinctions in statements about an event. The end result is measurements that may not accurately capture and represent the essence of expert military judgment.

The notion of requiring an expert to render precise statements about a complex military event appears to be incompatible. Zadeh (1973) proposed that a principle of incompatibility be applied in dealing with complex systems: "As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics" (p. 28). Given the complexity of military systems and operations, one can appreciate the principle of incompatibility. For example, being able to precisely document how many rounds were fired by a tank unit in a complex tactical engagement will most likely not reveal much about how that unit performed in the context of the whole battle. Here, the single objective indicant alone, although easily obtained, may have very little military significance. Focusing on objective dimensions of the battle tends to misdirect attention to physical event parameters that are themselves of little information value.

Furthermore, examining a multiplicity of such indicants together tends to induce information overload, creating confusion rather than insight. Clearly, some intelligent synthesis of the information is needed to form a meaningful pattern from a jumble of disconnected data. This can only be achieved by developing a systematic method of interpreting measures within a framework of military concepts and principles. Such a framework is used by the military expert to understand the meaning of battle events.

Background of the Problem

Military Performance Measurement

There are typically many dimensions of performance that can be assessed in any given military scenario. Many of these performance dimensions are influenced by the kinds of tactical requirements placed upon soldiers and their units, as well as the options for performance available to them. Because there are many levels on which to evaluate military performance, defining a comprehensive criterion that distinguishes successful performance from unsuccessful performance is often difficult.

Military performance measures tend to be complex in the sense that they contain objective physical and personnel data, and subjective judgment data. The multidimensional aspects of military performance becomes apparent when these data categories are considered simultaneously. Is a successful mission one in which the fewest rounds were fired and the least fuel consumed (physical data), the one whose units had the lowest casualties (personnel data), or the one which was rated as demonstrating a high quality tactical execution by a military expert (judgment data)? Clearly, all three aspects are important, but may mean different things in different situations.

The notion of properly defining performance criterion measures appears particularly important in the on-going military debate surrounding the issue of determining what device-based training strategies can actually accomplish. The issue of device-based training in general is a direct manifestation of new budgetary constraints on traditional training philosophies using operational equipment. Training predominately has been managed by the concepts of operating tempo (OPTEMPO) that determines fuel and maintenance costs, and live-fire gunnery exercises that set ammunnition costs. As a consequence of cost limits, training doctrine is becoming increasingly device-based rather than simply device supported (Burnside, 1990; U.S. Army Training and Doctrine Command, 1989).

However, as Burnside (1990) points out, "how should Army training managers face this dilemma of increasing the use of devices and simulations with only limited data available on what these tools will train"? Rendering device-based capability assessments is linked directly to problems that exist in defining measures of performance for device-based training systems. Prior to making recommendations about the types of military behaviors that can be effectively trained through simulation, one must first deal with the issue of developing performance measures that are based in some way on task standards essential for success in battle. Only then can intelligent assessments of device-based training be made in in terms of valid performance criteria.

The military has traditionally depended on SMEs who possess the domain of critical military concepts necessary for making complex performance-related judgments based on interpretations of objective data sources. For example, performance in device-based training simulations, as well as capabilities of the devices to train, commonly is assessd via judgments made by SMEs. In such instances, an SME observing an exercise might say that a platoon crossed the line of departure (LD) too early or too late, as a result of poor planning by the platoon leader. Here, the timing of LD crossing is described in relation to a prespecified time existing in an order, and linked to a prior cause. The SME's military concepts tell him how to abstract the difference between actual and ordered time to determine contextual meaning, and how to relate events causally. Both descriptive and comparative terms are used to interrelate different pieces of information so that a meaningful picture of a complex activity emerges. Within the context of this perspective, complex judgment can be thought of as an "emergent" feature of the interrelations between concept dependent terminology and objective data.

Although it is clear that value judgments by subject matter experts will continue to play a pivotal role in establishing performance guidlines and task standards in the military, it remains crucial that this method of assessment be continually subjected to tests of reliability and validity (Burnside, 1982). Much past research has illustrated the many problems associated with expert judgment. Biases that threaten reliability and validity come from many sources, including the context of judgment, personality, age, cognitive style, information processing limitations, judgment uncertainty and risk, stress and so on. As a result, researchers continue to examine the human's capacity for integrating diverse and partial information in rendering judgments, and what conditions alter the ability to judge accurately. Appendix A presents some alternative approaches and conceptual issues currently being considered by decision researchers in modeling both the nature of complex judgments and their underlying processes. The knowledge derived from this research will likley enhance our ability to develop measures of performance for device-based training programs that relate more closely to key military ideas and doctrine.

Automated and Instrumented Measurement

Although simulator combined arms training programs continue to evolve, future performance criteria are likely to be based, in part, upon developments like the Unit Performance Assessment System (UPAS). UPAS is a PC-based system that allows trainers to evaluate unit simulation performance. UPAS operates by collecting, from a variety of sources, real time data from networked interactive simulations, which include simulation networking systems for training (SIMNET-T), and research and development (SIMNET-D).

Briefly, SIMNET-T is a networked distributed processing battlefield simulator developed to complement combined arms field training exercises. SIMNET-T is located in the Combined Arms

Tactical Training Center at Fort Knox. SIMNET-D is a similar networked simulator that provides a reconfigurable test bed for prototyping futuristic weapons systems, organizations, and operational doctrine. SIMNET-D is in the Close Combat Test Bed facility, also at Fort Knox. These mannned simulator systems allow many players to engage in interactive, real-time battles against other human players or semi-automated forces locally or at remote locations in the U.S. and Europe. Data from these simulations can be entered into a relational database configured to resemble the National Training Center (NTC) database at Fort Irwin, California. The ARI Presidio of Monterey Field Unit maintains an archive of NTC databases for research purposes.

One objective being pursued in developing the UPAS system is to provide a low-cost capability to record many of the objective physical events that represent the critical elements of tactical missions and scenarios. In theory, UPAS should allow one to organize events characterizing a given scenario in a manner that is informative to trainers, and which can support training needs, analysis and research. For example, UPAS can replay vehicle movement and weapons firing events on a map display showing a bird's-eye view of the battlefield terrain. Magnified snapshots of the display at given points of a mission can be made from recorded event sequences that document key elements of a tactical mission. These snapshots show figures displayed over a terrain map providing detailed information on vehicle position, and gun tube and turret orientation. The replay and snapshot facilities of UPAS give trainers information to support evaluations of unit movement formations, coordination of actions, and execution of orders, as well as other features of the tactical operation.

Although UPAS should greatly improve the capacity to assess simulator-based training performance, developing performance measures from data collected on UPAS will be difficult. This will be especially true of complex performance measures that reflect more abstract functions, such as command and control.

<u>Classes of Events</u>. Implicit in many measurement systems is the notion of a hierarchical event structure which is used to categorize certain properties of a given phenomenon. The degree of specificity required, for example, to rank-order measurements of some phenomenon changes as the measures themselves become more general and fuzzy. For example, more global kinds of performance measures will be needed to address performance at the division level as opposed to the platoon level.

As one moves up the military echelon hierarchy, one begins to use more non-numeric response formats to communicate performance. This is essentially due to the fact that complexity makes it more difficult to make precise statements, because statements (or estimates) become conditioned by a multitude of other significant dimensions. For example, at a high level, such as a theater of operation, the measure may be one of "effective force structure". The measure of an effective force structure would tend to be linguistic rather than numeric. The measure would likely call for a value judgment which would combine data on the distribution of military resources, the immediate tactical situation, political ramifications, and so on.

However, measuring a more clearly defined event, such as "securing an objective on a terrain" would tend to move down the hierarchy to a lower level. Here, the performance measure generated from expert judgment would likely be a combination of non-numeric and numeric information. For example, pairing linguistic information (e.g., "good" execution) with numeric information (e.g., number of rounds fired, casualties taken, and positions occupied). Typically, the farther one travels down the echelon hierarchy to lower levels, the more the measures become increasingly numeric as in the case of firing accuracy and movement directions. The judgment of performance at lower echelons tends to facilitate higher degrees of precision in responses than the upper echelons simply because there is far less influencing the outcome of actions relevant to these levels.

In automated systems like SIMNET, there are computed events which are automatically recorded by the UPAS real-time data logger, such as elapsed time, vehicle location, status, and weapon firing that are obtained from the data broadcast over the computer networks. Similarly, at the NTC, instrumentation on vehicles collects and transmits information to telemetry stations for computer storage and processing. In both cases the recorded events are considered the primary elements of unit performance that are directly observable and very clearly specified. In addition, radio communications are directly monitored by observers and also can be recorded in analog form. A second order event class includes those events that typically can not be recorded or instrumented. This event class is not directly observable, rather, it is either deduced or inferred on the basis of directly observable events that have occurred.

The distinguishing feature of these particular event classes from that of even higher order measurement categories is that the events are typically considered to be binary in nature. For example, either the platoon crossed the line of departure at the ordered time or they did not. Here, the observation is made by reference to (i.e, mapped directly to) a physically instrumented event such as a marker code indicating passage of the line of departure. From a hierarchical perspective, the (psychological) distance between directly observable events and indirectly observable binary events is relatively small. That is, binary events can usually be reduced to their more elemental, directly observable parts. However, the relational simplicity between observable and nonobservable events ceases when discussing the more complex value judgments as the third class of events.

<u>Meaning of Events</u>. A fundamental problem of unit performance analysis lies in mapping observable events collected from UPAS or other sources onto the various theoretical elements

of training doctrine. Complex judgments made by experts of various military scenarios cannot typically be parsed into their most elementary physical events. This is because the referents of many judgments lie not just in the observable data, but rather in interpretations of the significance or value of certain tactical behaviors.

For example, if we are interested in assessing the mission of "movement to contact", UPAS will allow us to describe when and where all the tanks move and their spatial location in relation to terrain and each other. We can use this information to compute various indices of movement-to-contact performance such as unit speed, acceleration from the line of departure, and radial velocity of maneuvering vehicles. Many other measures are possible. However, no matter how much one quantifies what the tank units are doing in terms of measures generated from these kinds of observable events (i.e., elapsed time and position), the measures may not, by themselves, be informative. That is, the meaning we assign to given events is a function of an interpretation made within the framework of a set of military constructs. In any measurement problem you have observed indices of a phenomenon, and you have various relationships which tie these indices to a system of constructs and theories.

As data-based training devices become more complex and sophisticated, the link between computed or instrumented measures of performance (as recorded by UPAS) and the interpretations as to the success or failure of training formed on the basis of these measures may become more illusive and difficult to define. The future capability to alter and record many different parameters of a simulation is likely to make the job of performance evaluation more demanding. Using the current methodology for assessing training performance would likely include having experts generate value judgments on military behavior during or after the simulation. As simulations become more complex, the job of rendering judgments on performance will become more difficult. Complexity also will lead to more elaborate descriptions of military events by experts.

The problem of complexity can be ultimately defined as the relationship between the physical events recorded by simulation software and the interpretation of what these events mean vis-a-vis the domain of military propositional constructs. Hiller (1987) cogently makes a similar argument when characterizing the many shortfalls associated with performance measurement systems during field training operations:

> "Observers may intuitively feel that certain units are relatively effective or ineffective, but historically the training community has been unable to substantiate these feelings with hard, precise data. This drawback is somewhat analogous to the measurement problem in physics commonly referred to the Heisenberg Uncertainty Principle. Its three premises are that the process of measurement dynamically

affects the object being measured, that the object has many different potential states of existence, and that the object is known only (emphasis added) though measurement."

Hiller makes the points that evaluating unit performance can be an uncertain enterprise, and that, in principle, it is similar to the measurement uncertainty found in high energy physics: (a) that a tank unit will perform differently under the watchful eyes of expert military observers; (b) that the state of the unit itself is always in flux due to changes in personnel resulting from turnover, casualties, etc.; and (c) that performance assessment must be based exclusively on "snapshots" during training exercises because of the difficulty of performance assessment in the "home-station environment". The last point, which relates to the last Uncertainty Principle premise (i.e., the object is known only through measurement), can also be represented in the manner discussed in this report. That is, higher level measurements cannot usually be reduced to their elementary physical events. Instead, elementary events form the basis of some non-invariant linguistic transformations that are used in producing value judgments. These transformations are (tactically) context and construct dependent. Thus, they exist only as complex linguistic transformations.

What the above discussion by Hiller suggests is that if performance measures at all levels of military echelons are to be useful, they must be used intelligently and must be able to generate reliability in outcomes. This can only happen in the framework of a theory for developing and using such measures. Within the context of searching for a theoretical framework that can support device-based performance measurement procedures, two observations are apparent:

- There is a conspicuous absence of a universally accepted taxonomy for performance measures that should be used in device-based training programs.

- Of the measures that currently exist, there appears to be little in the way of standard mathematical definitions which characterize how the measures can be combined.

In a recent compliation of Army-related measures of effectiveness (MOEs), Feng (1991) grouped measures by system functions closely related to the seven battle operating sytems (BOS) used by the Training and Doctrine Command to organize miltary studies, system analyses, and operational tests (U.S. Army Training and Doctrine Command, 1990). While commonly used, the BOS are high level classifications that relate imperfectly to the lower-level tasks used to assess unit performance. Many tasks can be found to clearly contribute to more than one BOS.

Furthermore, while Feng listed numerous measures in each category, there was no specific guidance indicating appropiate circumstances for their use. As Feng notes:

"The measures have been culled from various sources but should not be considered a definitive list by any means. Rather they are offered as examples of what one might use in tests and studies. Better or more appropriate MOE's should always be developed wherever possible."

The absence of standard mathematical definitions for high-level concepts means that the measurement process must rely on informed, but often arbitrary and controversial procedures for creating complex measures of effectiveness and performance. Accepted, validated rules for combining measures do not exist. The consequences of working with arbitrary measures renders causal conclusions about performance suspect. Currently, at all levels of military echelon, objective measures, as well as complex value judgments, are treated simply as elements in the field of real numbers that are assumed to reflect aspects of the causal process under investigation.

<u>Fuzzy Set Theory</u>

Linguists have long acknowledged that people understand and operate on natural language concepts. However, in most performance measurement systems there is a rigid standard enforced that eliminates all vague information in favor of information that is extremely precise in nature. This rigid adherence to precision significantly reduces the ability to discover fundamental conceptual functions (Zadeh, 1973). Zadeh (1973) has developed quantitative techniques for dealing with the vagueness in natural language. The techniques are based on fuzzy set theory, which represents an extension of the traditional theory of sets.

The unique feature associated with fuzzy logic is that it permits a complex system to contain both numeric and linguistic variables, where the linguistic variable is a label for a fuzzy set. Fundamental to fuzzy set theory is the notion of using a linguistic variable as a means of estimating the possibility of an event being a member of a given fuzzy set. The power in using a natural language approach to estimation lies in the ability to provide a method in which to model the often imprecise activities associated with military operations in a manner that closely parallels how military personnel think about these activities. Rendering estimates or predictions of complex military phenomena is based largely on notions of judgment, best guesses, intuition, and having a good feel for the battlefield. In addition, the experts, who are assumed to possess these somewhat vague attributes and abilities, clearly differ as a result of the differences in the breadth and complexity of the military construct knowledge that each draws upon to make such judgments, best guesses, etc.

Linguistic variables differ from a numerical variable in the sense that their values are not numbers, but, rather words or phrases of a natural language (e.g., English). Here, words are used to communicate quantity and magnitude information, however, in a manner that reflects the imprecision of a given complex and ill-defined problem. For example, the linguistic variable "distance" may take on the values of "very far", "rather far" "far", "not very far", "somewhat close", "close", "very close" and so on. The assumption underlying these fuzzy sets is that the transition from membership to nonmembership is a gradual one, and not a step function. This contrasts with nonfuzzy set theory where a membership function precisely indicates what elements are members of a given set, and what elements are not.

Fuzzy sets then represent restrictions on the values of a given linguistic variable. Figure 1 shows some characteristics of restrictions on the linguistic variable, distance. The figure could represent a linguistic example of distance estimates encountered in an "indirect fire" artillery situation. As indicated in the figure, as one moves down the distance axis toward greater and greater values of the distance variable, one transitions from one linguistic restriction to the next.

The degree to which a distance value belongs to a particular linguistic set (i.e., its restriction) is determined by the numerical value which characterizes the possibility or plausibility of a given value belonging to a given set. The possibility or membership values over the whole range of the variable are given by the membership function for the variable. Thus, the values of the membership functions shown in Figure 3 indicate that the distance value example of 250 meters belongs to the set "Very Close" more than to the set "Close". Similarly, the distance value 1050 meters belongs more to the set "Far" than "Very Far".

As was indicated above, fuzziness is entirely distinct from the concept of uncertainty in probability. The uncertainty associated with obtaining a particular value of a roll of a die has a particular probability. There is no vagueness involved in the problem--just a lack of knowledge concerning a given future event. However, once this knowledge becomes available, the problem is completely determined. In contrast, when dealing with the issue of vagueness, no matter what one does, a concept will apply more to some elements than to others. That is, no matter how much information you have on the fuzzy variable "distance", the boundary between "far" and "not far" will be imprecise.

Membership in fuzzy sets is specified in the same manner as nonfuzzy sets: by roster, a relation, or an algorithm that defines a function mapping elements from a universal set to the fuzzy set in question. This mapping generates values for every element in the universal set, such that each element is paired with a numeric quantity in the closed interval [0,1] indicating its grade of membership in that fuzzy set. Once this membership function has been defined, the set can be used as a linguistic variable in fuzzy inferences and algorithms can be manipulated by set theory operations such as union and negation.

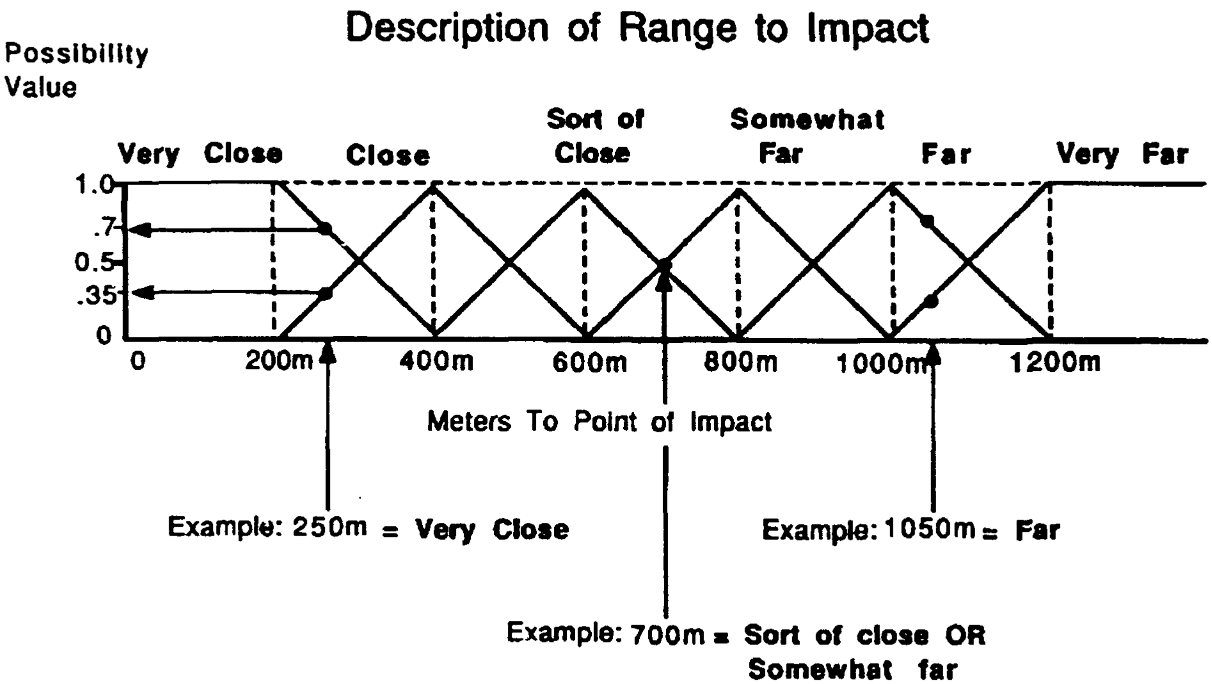

Figure 1. Fuzzy linguistic restrictions on a distance variable.

The operators used with fuzzy sets are extensions of similar operators used with nonfuzzy sets. Essentially, negation, logical and algebraic operations, hedges, and other terms that modify the representation of linguistic variables can be considered labels for various operators defined on fuzzy subsets of a universe X. Some of the fundamental operators are described here. For a more complete overview of basic operations see Schmucker (1984), Smithson (1987), and Zadeh (1973).

Let the function defining degrees of membership in the set A be denoted by $f_A(x)$, where $x \in X$. Then the set A is defined as:

$$A = \{ x \mid f_A(x) > 0,\ x \in X \}. \qquad 1.0$$

Similar to nonfuzzy set theory, the complement operation corresponds to negation. The complement of set A ("not A") is denoted "A'" and is defined as:

$$A' = \{ x \mid [1 - f_A(x)] > 0,\ x \in X \}. \qquad 2.0$$

The fuzzy union of two sets is analogous to the inclusive "or" operation in nonfuzzy set theory. The union of two fuzzy sets A and B is denoted by $A \cup B$, and is defined as

$$A \cup B = \{ x \mid [f_A(x) \vee f_B(x)] > 0,\ x \in X \} \qquad 3.0$$

where $$f_A(x) \vee f_B(x) = \max[f_A(x),\ f_B(x)]. \qquad 4.0$$

The fuzzy intersection is the analog of the nonfuzzy set "and" operator. The intersection of two fuzzy sets A and B is denoted as $A \cap B$, and is defined as:

$$A \cap B = \{ x \mid [f_A(x) \wedge f_B(x)] > 0, x \in X \} \quad 5.0$$

where

$$f_A(x) \wedge f_B(x) = \min[f_A(x), f_B(x)]. \quad 6.0$$

The above description serves to illustrate how various linguistic operators can be defined in terms of fuzzy sets. All of these operators reduce to their corresponding nonfuzzy set operators when f(x) is binary, i.e., is limited to only values of 0 or 1. Smithson (1987) demonstrates that other linguistic operators (e.g., very, somewhat, and other terms usually without unfuzzy analogs) can be incorporated into the theory of fuzzy sets by being defined as specific operators on membership functions. For example, given the fuzzy set labeled A, and denoting "very" by "+", "very A" should be of the form:

$$+A = \{ x \mid f_A^a(x) > 0, x \in X, a > 1.0 \}. \quad 7.0$$

Similarly, given the definition of "not" (Equation 2.0) and "very" (Equation 7.0), Zadeh (1973) and others felt that "not very" should take the form:

$$(+A)' = \{ x \mid [1 - f_A^a(x)] > 0, x \in X, a > 1.0 \}. \quad 8.0$$

Appendix B describes some of the rationale for an experimental approach designed to examine whether military personnel actually utilize language terms according to the fuzzy theory operations illustrated here and in Smithson (1987). The research outlined in Appendix B represents initial efforts to test the results of various operations on a group of natural language military terms in order to verify some basic fuzzy set transformations.

Case Study of a Domain of Military Concepts

Because this paper represents a discussion of the requirements for moving toward a fuzzy theory of device-based performance measures, a proposal for a method of obtaining a definition of a performance measure is advanced in operational terms.

In order to reach the definition of a measure, we must, in this first phase, construct the process leading to the selection of a candidate measure. There are several steps that can be included in the process of developing a useful semantic network that can provide a framework for interpreting military constructs:

- Identify a set of military propositions.

- Identify the semantic network underlying interpretation of the military propositions.

- Rank order the important language elements that quantify the primary military propositions as a step toward computing the (fuzzy) truth values of natural language propositions.

- Employ a methodology for obtaining estimates of membership values to complete the quantification. The extension of a military proposition then is identified by a corresponding fuzzy set with membership function that indexes the truth value of the proposition when applied to specific cases.

Our aim for the first phase of developing a fuzzy theory of natural language to support the measurement process in device-based training is to select for study a key domain of military propositions. Once we have the domain, then we need to elicit the language characteristics used by experts in applying the propositions within that domain to military observations or data.

Selecting Natural Language Expressions

Clearly, one key to adequately representing fuzzy restrictions on a military concept lies in selecting the appropriate set of natural language expressions. The expressions will serve as values for the linguistic variable chosen to capture the military construct of interest. Since we are essentially dealing with a natural language approach to scaling fuzzy restrictions on a linguistic variable, we need terms that play the role of language elements. Although, many terms can be used that represent the various elements of language syntax, terms that play the roles of "primary" and "hedge" are most useful (Schmucker, 1984). Primary terms are usually adjectives (often adjectives of degree or comparatives), while hedges are adjective modifiers (often intensifiers). Combinations of primaries and hedges may also be joined in range or relational phrases by terms such as "to", "and", or "or". Table 2 shows a sample list of natural language expressions commonly used in risk analysis that illustrate some of the possibilities.

Table 2

Examples of Natural Language Expressions[1]

High	Low
Medium	Not High
More or Less High	Medium to Sort of High
Indeed Low	Slightly Lower than Pretty High
About 4 to about 6	Not Higher than Medium
Higher than Low and Lower than Sort of High	

[1]Expressions taken from Schmucker, 1984.

The essential goal in developing a usable set of expressions that can eventually be scaled in order to compute truth values of natural (i.e., military) language propositions would be to: (a) identify a set of relevant primary terms that would serve as "adjectives" in a natural language grammar, (b) identify a set of hedges that would serve as intensifiers moderating the various adjectives, (c) identify a group of simple phrases that would combine hedges and primary terms, (d) determine if a set of relational phrases or compound phrases could be included in the set of expressions used to restrict a linguistic variable.

The domain associated with those military constructs that address and describe military communication has been chosen for this phase of the project. Communication seems to lend itself well to fuzzy representation. Commanders tend to use natural language responses in grading communication performance (Babbit & Nystrom, 1989). In addition, the concepts that are central to communication (e.g., situational context, message content, timing of reports) are themselves fuzzy entities. It is difficult to perceive, for example, a given spot report to be either a member of the set "those reports having message content", or not. Instead, it is more believable to consider the transition from membership to nonmembership as gradual as opposed to abrupt. So it is likely that the essential attributes associated with military communications are graded concepts, and that reports differ in degrees of timeliness and message content. Appendix C[1] (Bessemer, 1991a) identifies a number of concepts commonly used to describe reporting performance, and presents hypotheses about fuzzy relations among these concepts.

Because there tends to be a vast number of possible language expressions which can represent the linguistic values of a primary term, rules have been formulated that guide the selection of these expressions. One particular rule-based approach that is frequently applied by computer scientists has been called Backus-Naur Form (BNF, Schmucker, 1984). BNF specifies a series of linguistic categories which contain language elements necessary for the flexible manipulation of language concepts. BNF notation specifies those linguistic terms which would logically fit, for example, a rating category, a range phrase category, a hedged phrase category, and so on. The notation essentially provides a rule for selecting linguistic expressions that represent a well conceived and flexible group of language terms that can be used in linguistic description.

Because of the exploratory nature of this work, a rule-based scheme was not carefully implemented for soliciting expressions. The objective here was to elicit from the commanders the language terms they felt comfortable in using to describe, and linguistically quantify, various aspects of communication. For a summary introduction to BNF notation, see Schmucker (1984).

---

[1]Reproduced here with permission of the author.

Interview Method

The methodology used here to identify a semantic network that is useful for interpreting the performance of a sender's communication skills was called progressive elaboration. The objective in using this unstructured interview approach was to have a group of subject matter experts delineate the levels of a semantic network for assessing communication performance.

Subjects. Six company commanders from the 194th Armored Brigade, Task Force 1-10 Cavalry, served as subject matter experts for the first phase of the project. Task Force 1-10 Cavalry is an active army unit. The company commanders were experienced in conducting force-on-force unit training at Fort Knox, Kentucky.

Procedure. The six commanders were field interviewed in two groups of three on two different days. During the interview, the three commanders were briefed on the nature of the project and the research goal of identifying how the commanders trained the platoon leaders on various aspects of communications. The commanders were asked a series of training oriented questions about various aspects of military battlefield communications. The first question posed to the commanders asked how they provided feedback on a platoon leader's performance at sending reports. The commanders answered the question by outlining the context in which the importance of communication changes, and basic standard operating procedures employed in tactical operations. The commanders were then asked to elaborate futher on topics raised in their initial answers. When descriptive terms occured in the answers, follow-up questions asked for associated terms and relations between terms.

It was interesting to note that all of the commanders seemed to agree that communication performance was very difficult to quantify and evaluate precisely. However, the commanders agreed that communication was always a fundamental precursor to military engagements, and that communications, in large part, determined the successful outcome of a mission.

Results of Interviews

Semantic Networks. Over the course of the interviews, the commanders identified three basic constructs that they felt were key in assessing the performance of a platoon leader's situation and spot report communications. The three constructs were essentially viewed by commanders as being linked to: (a) the necessity of reports, (b) the timeliness of reports, and (c) the informativeness of reports. Although, these constructs that were the primary focus of the interviews, it is important to realize the context in which they were judged. For example, the relationship between the three constructs must be considered within the framework of report type and format. Relationships among the constructs are schematically summarized in Figure 2.

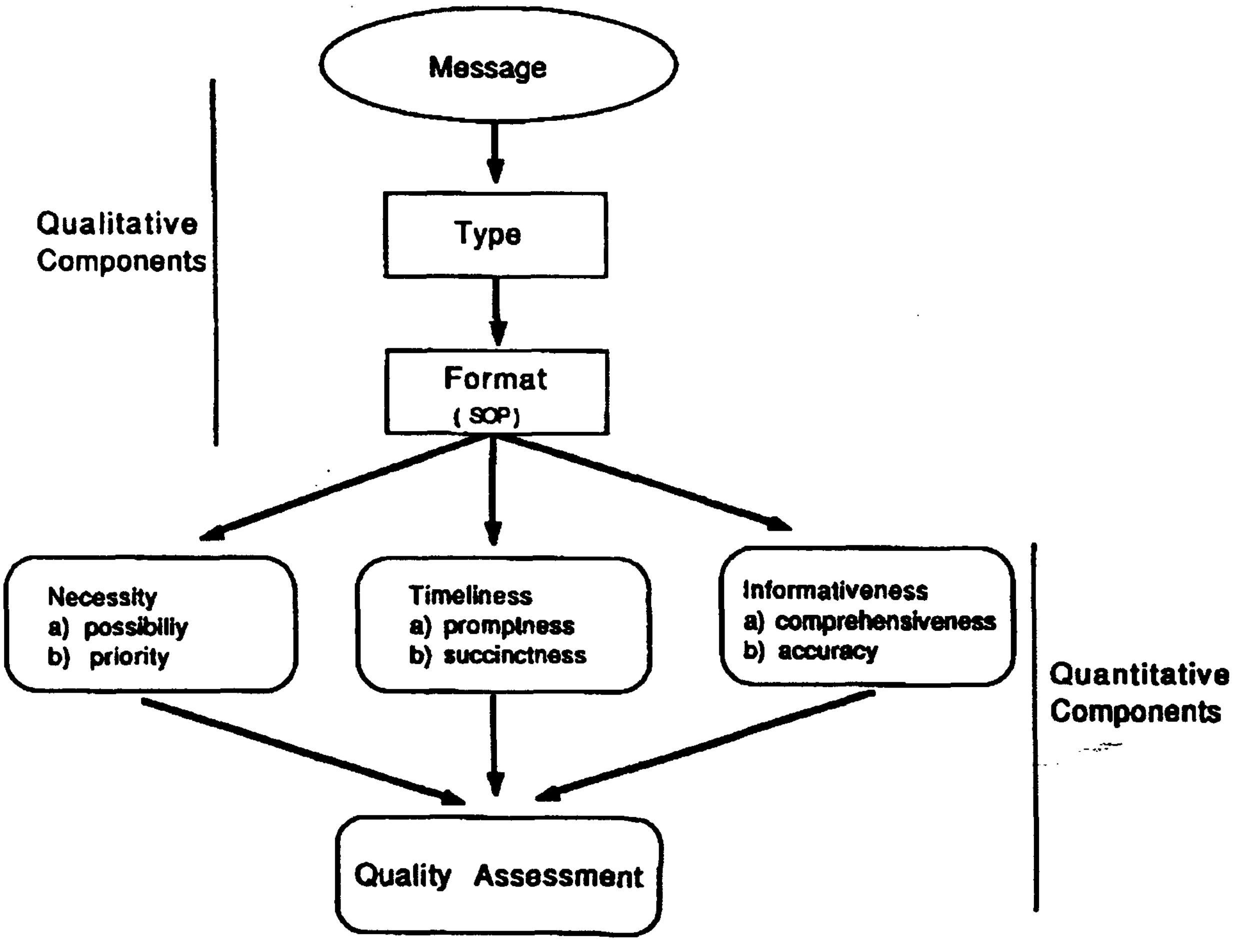

Figure 2. Quality assessment hierarchy for battlefield reports.

The relationship between the various constructs can be conceptualized as a flow chart representing the hierarchical structure of a communication assessment system. The system characterizes the universe of discourse, which in this case is partitioned into qualitative and quantitative component dimensions.

The qualitative dimension of the system essentially defines the messages by type and format. In this case, the type of report would refer to the kind being sent. A number of report types are listed in Appendix C, Table C-1. Format refers to the structure of the report content actually sent over the radio network. The format of a report is dictated largly by standard operating procedures (SOP) for report transmission. These elements form linkages with the quantitative dimension of the system, which itself defines the constructs of communication that serve as continua for ranking the quality of messages. Here, the qualitative characteristics of a communication report influences the assessment process. Thus, overall quality assessment of a given report will depend on the nature of the report and its format.

In the interviews, the commanders tended to describe the notion of necessity in Figure 2 as being related to the issue of whether a report was required to be transmitted in a given tactical situation. This judgment was made based on aspects of the situational context in which the report was generated. The commanders indicated that new platoon leaders tend to avoid any communication with other units and with their commander. However, they felt that this was a transient phenomenon linked primarily to the novel, and perhaps, stressful experience of acting as platoon leaders in a force on force encounter. In contrast, the commanders indicated that once the initial shock of acting as platoon leaders passed, unnecessary communication tended to be a significant problem on the battlefield.

Within the necessity dimension were subordinate elements that apparently modified certain properties of the construct. One point that emerged from discussions with commanders was that necessity was related to a platoon leader actively acquiring relevant information for transmission as a report. The inference made here is that a "necessary" communication must be "possible" to the extent that the information needed to make the report is available to the platoon leader.

A second subordinate element in the necessity communication construct described by commanders was related to the issue of a platoon leader's ability to successfully prioritize his activities. Here, priority refers to the ability to organize activities on some importance dimension in order to capitalize on, perhaps, a lull in the battle, which would permit the time needed for sending a report. At issue is the notion of judging the point in time where the platoon leader feels that sending a message becomes of prime concern, and then organizing activities around this goal.

Another dimension, however, to the priority concept concerns a platoon leader's judgment about his report in relationship to the total battlefield communication activities occurring at any given time. During the interviews, the commanders explained that because communication networks can accommodate only a finite amount of radio traffic, platoon leaders must exercise discretion in evaluating message priority on the basis of the information being transmitted by other platoons. In other words, a necessary message is not only possible (i.e., the information is available and has been acquired), it must be given a priority judgment rating both in terms of: (a) its importance in the scheme of activities that must be performed by a platoon leader and (b) how much room on the communication network exists for transmission. This latter priority dimension implies that a platoon leader must evaluate, to some degree, the importance of his message in light of: (a) a fixed amount of network space and (b) the importance of communications being transmitted by other platoons.

Although there appeared to be significant individual differences associated with the language the commanders used to grade the necessity component of communication, as well as other dimensions of communication, there was some indication that the subcomponents tended to be discrete in nature. While commanders tended to speak in degrees of necessity, the notions of "possibility" and "priority" were viewed as either/or situations. Furthermore, the commanders described having the ability to recognize, during a battle situation, a threshold where relevant information became available for transmission as a report, and similarly, the point at which a platoon leader exercised good judgment in prioritizing his activities.

The primary terms used by commanders to characterize necessity of communication are presented in Table 3. The terms, as well as the commander's descriptions of communication, were recorded from written transcripts of interviews with two groups of commanding officers. After the interviews were concluded, all the primary and hedge terms were identified in the transcript. With the exception of the terms "trivial", "minor", and perhaps "significant", the primary terms used by the commanders to define necessity appear to connote some sense of urgency. However, it was unclear from the discussions with the commanders how these terms differed with respect to what the commanders had in mind when asked about the necessity dimension.

Table 3

Primary Terms Used to Grade Necessity of Communication

Critical	Serious	Trivial	Pivotal
Important	Significant	Minor	Dangerous

The network of expressions that commanders used in conjunction with the primary terms in Table 3, providing a means to convey different degrees of the terms, were strikingly similar to one another. That is, although certain commanders preferred to use particular primary terms in describing "necessity", they each used similar expressions in order to linguistically quantify the construct. The fact that few expressions were recorded from discussions with the commanders may indicate that the necessity concept is not as vague as, perhaps, other military concepts. For example, no range phrases of the sort "very critical to critical" were recorded from commanders, nor were any of these more complex phrases used to describe the other communication constructs. Table 4 presents a list of the expressions used to restrict the meaning of the primary terms shown in Table 3. Both the primary terms and restrictive expressions listed here appear to comprise, at least in part, the linguistic categories needed to form some of the more basic grammatical elements of English language. Here, it is apparent that the primary terms and their restrictive expressions play roles somewhat analogous to that of "adjective" and "modifier" in the construction of English language phrases.

Table 4

Linguistic Expressions for the Primary Terms Associated with the Concept of Necessity

Very	Somewhat	Extremely	Particularly
Not Very	Nearly	Possibly	Absolutely
Likely	Unlikely	Not Likely	Rather

A second concept that commanders viewed as being essentially linked to tactical communications was the timeliness of reports by platoon leaders. The commanders expressed the need to have information concerning battlefield situations as soon as possible. Furthermore, in the larger context of total mission requirements, the timing of reports would ultimately influence the nature of tactical strategy and operations.

Evaluating the timeliness of a platoon leader's reports was considered to demand assessing at least two other subelements of the timeliness construct. The first element was viewed as being related to the idea of direct and firsthand reporting of battlefield events. In querying the commanders more closely, it seemed the idea of direct and firsthand reports could be reduced to the notion of promptness in reporting activities. Thus, the report was viewed, in part, as a report that was made as quickly as possible after acquiring the information.

The promptness element of the timeliness construct was complemented by the notion of a platoon leader sending a concisely organized report. Here the commanders felt that a good report was also succinct in the sense that it was organized according to standard operating procedures, and was delivered without error or unnecessary interruptions and/or delays. Thus, the idea of timeliness could be partitioned by the notions of promptness and succinctness. The commanders further indicated that the notion of succinctness would depend, in part, on the kind of report being sent. This latter qualification to the succinctness dimension was due to the fact that reports would vary in length depending on the nature of the report (e.g., situation report, shell report, etc.). Both of these subcomponents were described by commanders as being assessed in a discrete (i.e., either/or) manner. So while the commanders clearly described timeliness in a graded fashion, the subcomponents appeared to be viewed as dichotomous in nature.

Table 5 presents a list of primary terms that commanders used when discussing the concept of timeliness of reports. The commanders evidently viewed the timeliness concept as existing on a "goodness" continuum of sorts. Further, the commanders seemed to encapsulate the entire timeliness concept as an effort in "timing" of reports. The notion of report "timing" versus "timeliness" connotes somewhat different, albeit, related ideas.

Table 5

Primary Terms Used to Grade Timeliness of Reports

Superior	Outstanding	Good
Average	Adequate	Acceptable
Moderate	Ok	Poor

In this case, the way in which commanders described the concept may have been different from the way they considered the concept in practice. Table 6 lists expressions that commanders used as sets of restrictions for the primary terms given in Table 5.

A final aspect of communication that was viewed by commanders as being instrumental to the overall quality of reporting activity was related to the information value of messages. However, the informativeness of reports was seen as being more difficult to evaluate than the concepts of necessity and timeliness. This difficulty was due, in part, to being unable to restrict the evaluation to a single platoon leader's performance. While the commanders agreed that a key dimension of communication was its information value, they indicated that this was a very context dependent construct. That is, the information value of a given report would be dependent on the battlefield situation at the time of the report.

However, the commanders did indicate that evaluating the information value of a platoon leader's communication, from the point of view of the sender, could be made on the basis of completeness and precision. This is to say, that sending a complete report would be more informative, all things being equal, than an incomplete report. Here, the commanders noted that a formatted report (e.g., spot report, situation report) could be judged on the comprehensiveness dimension because critical report elements were outlined in the SOP for a given class of report. Thus, the comprehensive report was seen as meeting the criteria outlined in current tactical doctrine, which itself was defined in the standard operating procedures established for the kind of report being sent.

The commanders also indicated that it was more difficult to apply a standard evaluation procedure for unformatted reports, such as ones communicating position and movement information. Commanders appear to use the SOP as a benchmark for assessing completeness of reporting activities. Without an SOP to guide evaluation, the completeness dimension becomes less defined.

The second element of informativeness was viewed by commanders as having to do with the precision or accuracy of a platoon leader's reports. Informativeness was considered not only a function of the completeness of a report but also its validity. An example of an inaccuracy would be sending a

Table 6

Primary Terms with Linguistic Expressions for Timeliness of Reports

Superior	Good	Average
a) Extremely	a) Very	a) Somewhat
b) Very	b) Somewhat Very	b) Very
c) Somewhat	c) Pretty	c) About
d) Rather	d) Quite	d) Not Very
	e) Fairly	e) Not
Outstanding	f) Good in Most Cases	
a) Very	g) Not	OK
b) Really Very	h) Not Very	a) Somewhat
	i) Not Quite Very	
Poor		Adequate
a) Very	Acceptable	a) Fairly
b) Somewhat Very	a) So-So	b) Not
c) Pretty	b) Barely	
d) Not		Moderate
e) Not Very		a) Somewhat

situation report indicating the position of enemy vehicles when the true identity of the vehicles was friendly.

However, it was also apparent that the accuracy element of the construct informativeness tended to be viewed by commanders as discrete in nature, much like the subelements of the constructs necessity and timeliness. Commanders gave examples, such as in the case of a shell report, where all the essential elements of the report are accurate with the exception of the grid location of enemy artillery. Here, the report was considered important and informative in the sense that it identified a significant threat to the company. However, it failed to locate the threat accurately, and as a result would be viewed as partially informative. Table 7 lists the primary terms along with their linguistic values used by the commanders when describing the information value of a particular report.

Although the network of terms used to characterize military communication tended to be fairly broad, the adjectival (primary) terms and modifier (hedge) terms tended to be well defined and few in number. These terms reflected a reasonably rich set of expressions that seemed to afford flexibility in subjective assessments of performance. However, there was a significant amount of individual differences associated with the terms and expressions used by the commanders to grade the various dimensions of communication. Therefore, in an effort to better understand what terms and expressions were viewed by commanders as most important in describing communication in general, commanders rank-ordered the terms on the degree to which they belonged to the basic language associated with describing

Table 7

Primary Terms with Linguistic Expressions For Informativeness of Reports

High	Moderate	Average	Low
a) Very	a) Very	a) About	a) Very
b) Somewhat	b) About	b) Somewhat	b) Somewhat
c) Fairly	c) Fairly	c) Very	c) Fairly
d) Not	d) Right About		d) Not Very
e) Not Very			e) Rather
f) Really Not Very			

military communication. The commanders were asked to sort the terms into three categories: (a) those terms showing a high affinity for describing communication performance, (b) those showing a moderate affinity, and (c) those showing little or low affinity for the language of communication.

The results of that ranking process are shown in Table 8. Apparently those terms which characterize both degrees of good and bad, and degrees of excellence, were viewed as being highly related to the descriptive domain of military communications. On the other hand, the terms that reflect degrees of acceptability appeared to define what was considered a group demonstrating a moderate affinity for communication language. Finally, several terms were ranked as showing little relationship with the language of military communication.

One might argue that the categories differ on a dimension of precision. The high affinity terms foster clear statements about communication. In this respect, the terms manifest a lower sense of vagueness than the second and third term categories. However, it may be possible that when the context in which communication takes place becomes more complex and uncertain, commanders may use a broader array of linguistic terms in order to better describe the various features of communication.

Table 9 shows the results of commander's rank-ordering of the High Affinity group descriptive terms on the basis of the merit in the terms for reflecting "goodness". Although, this ranking approach does not lead to the scaling of terms along some continuum, it does provide a means for establishing the relative degree of goodness for the terms. Surprisingly, there was complete unanimity among the commanders on the rank ordering of the terms. Once again, this may have been due, in part, to the fact that the High Affinity terms tended to be associated with very precise meanings. Babbitt and Nystrom (1989) have noted that the precision in terminology is inversely related to the amount of variance in peoples' responses to terms. Here, the commanders seem to have little difficulty determining whether one term denoted a higher degree of goodness/excellence than another.

Table 8

Term Categories Based on the Affinity for Describing Communication on a "Goodness" Dimension

High Affinity	Moderate Affinity	Low Affinity
Good	Acceptable	Normal
Very Good	Fairly Acceptable	Important
Extremely Good	Highly Unacceptable	Barely Adequate
Outstanding	Somewhat Average	So-So
Quite Good	Good in Most Cases	Very Important
Superior	About Average	Minor
Somewhat Poor	Trivial	Rather
Very Bad	Fairly OK	Dangerous
Poor	Fairly Good	
Not Good Enough	Significant	

Table 9

Rank Ordering of High Affinity Terms on a "Goodness" Dimension

1.	Outstanding	6.	Good
2.	Superior	7.	Not Good Enough
3.	Extremely Good	8.	Somewhat Poor
4.	Very Good	9.	Poor
5.	Quite Good	10.	Bad

Discussion

The use of various terms by commanders that favor precise meanings might be argued a manifestation of discipline in military training as well as the possible linguistic constraints imposed by the domain of military communication propositions. Soldiers are typically reinforced for being precise, succinct and clear when interacting with commanding officers. The tendency to "waffle" when communicating subjective assessments of various military operations may be viewed by commanders as shrinking from responsibility for one's judgments and decisions. Furthermore, language phrases containing vagueness or uncertainty can be interpreted by superior officers as reflecting a lack of confidence and/or knowledge regarding a particular subject area. In this case, a superior officer may attribute imprecision in expressing an assessment of some military situation to a shortcoming in the soldier, rather than to an obscure situation.

Evidence for a reluctance to use uncertain terminology comes from the complete absence of range and other complex phrases in describing aspects of communication behavior. Range phrases play an important role in facilitating flexibility of natural language expression in other linguistic domains such as risk analysis, human performance modeling, and occupational safety). However, for the soldier, responding with the phrase "good to very good" may foster negative impressions of the soldier rather than communicate the ambiguity of the phenomenon being judged.

However, an alternate hypothetical possibility may be that the military scenarios that the commanders drew from memory as exanples considered during the interview process were fairly well bounded. That is, they had recently observed relatively standard military operations that may have been rather easy to assess. In this sense, the use of a smaller set of linguistic terms may have been due to the simple fact that the exercises were easy to observe and not overly complex. It would clearly be interesting to explore the possibly of the commanders using a much richer set of language terms under more difficult and less well defined circumstances. This is to say that, given complex and vague conditions, it is possible that commanders may utilize a greater variety of terms to quantify performance because there are simply more degrees of freedom associated with grading that performance.

The primary findings in this case study, however, point to the notion that there is a fairly well defined set of military constructs that are associated with the acts of forming and sending communication reports. The commanders were all adamant that reports should be assessed along the three dimensions presented in the results. While some uncertainty remains about how the commanders actually make multidimensional judgments of each primary communication construct using the subcomponent information (e.g., possibility and priority for the construct timeliness), it is clear that the primary constructs are thought of in degrees of quality. Thus, it should be possible to model these primary terms with fuzzy set theory techniques. Appendix B outlines a rationale and approach to studying how commanders use and combine subcomponents found in the case study as they come to judgments about the quality of primary constructs.

Summary

This report represents a first step in developing a performance measurement system that is based on fuzzy set theory. It outlined the logic associated with making complex judgments of performance, highlighting the notion that judgments contain conceptually-based linguistically-expressed interpretations of events that are embedded within a framework of military theory. The report briefly summarized the serious restrictions that are placed on judgments by a performance measurement process that imposes an artificial precision on how these measures can be represented. The argument advanced in this report is that the rigidity of such measurement procedures both conceals and obscures expert military judgment by disallowing the imprecision associated with the natural language that ties physical events to a framework of military ideas.

The report offers an alternative to traditional performance evaluation methods that allows modeling complex behavioral systems that contain both numeric and linguistic variables. Fuzzy set theory is presented as a formal method for modeling the natural language expressions that form the basis of value judgments made by military experts. Further, it is suggested that fuzzy set theory may be useful in connecting instrumented physical measures of combined arms simulator training with the subjective measures of expert military observers.

Finally, the report documents a case study of a military domain of constructs wherein a sample of commanders identify three dimensions of communication and the semantic networks used to quantify these dimensions. The networks are summarized as a possible set of terms that can be used to document the validity of fuzzy set operations in predicting how commanders manipulate and use the terms in quantifying communication performance.

References

Arbib, M. A. (1972). The metaphorical brain. New York: John Wiley & Sons.

Babbitt, B. A., & Nystrom, C. O. (1989). Questionnaire construction manual (Research Report 89-20). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A212 365).

Bessemer, D. W. (1991a). Relations among concepts describing reports. Unpublished manuscript.

Bessemer, D. W. (1991b). Personal Communication.

Brehmer, B. (1973). Effects of task predictability and cue validity in the learning of probabilistic learning tasks. Organizational Behavioral and Human Performance, 11, 1-27.

Brehmer, B. (1976). Subjects' ability to use functional rules. Psychonomic Science, 24, 259-260.

Brunswik, E. (1952). Conceptual framework of psychology. Chicago: University of Chicago Press.

Budescu, D. V., & Wallsten, T. S. (1979). A note on functional measurement and analysis of variance. Bulletin of the Psychonomic Society, 14, 307-310.

Bung, M. (1980). The mind body problem. Oxford: Pergamon Press.

Burnside, B. L. (1982). Subjective appraisal as a feedback tool (Technical Report 604). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A138 873).

Burnside, B. L. (1990). Assessing the capabilities of training simulations: A method and simulation networking (SIMNET) application (Research Report 1565). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A226 354).

Dawes, R., & Corrigan, B. (1979). Linear models in decision making. Psychological Bulletin, 81, 95-106.

Edwards, W. (1977). How to use multiattribute utility measures for social decision making. IEEE Transactions on Systems, Man and Cybernetics, SMC-7, 326-340.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Baysian statistical inference for psychological research. Psychological Review, 70, 193-242.

Feng, T. (1991). Measures of effectiveness compendium (Research Product 91-07). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 377)

Guilford, J. P. (1975). Factors and factors of personality. Psychological Bulletin, 82, 802-814.

Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswick's integration of the history, theory and method of psychology. In K. R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston.

Hammond, K. R., Stewart, T. R., Brehmer, B. & Steinmann, D. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes, New York: Academic Press.

Hiller, J. C. (1987). Deriving useful lessons from combat simulations. Defense Management Journal, 2nd & 3rd Qtr, 29-33.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. Management Science, 31, 680-702.

Lakoff, G. (1973). Hedges: A study of meaning criteria and the logic of fuzzy concepts. Journal of Philosophical Logic, 2, 458-508.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Polkinghorne, D. E. (1984). Further extensions of methodological diversity for counseling psychology. Journal of Counseling Psychology, 31, 416-429.

Schmucker, K. (1984). Fuzzy sets, natural languages and risk analysis. New York: Computer Science Press.

Simon, H. A. (1978). Rationality as a process and product of thought. American Economic Review, 68, 1-16.

Slovic, P., Fischhoff, & Lichtenstein, S. (1977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Baysian and regression approaches to the study of information processing in Judgment. Organizational Behavior and Human Performance, 6, 649-744.

Smithson, M. (1987). Fuzzy set analysis for the behavioral sciences. New York: Springer-Verlag.

Tucker, L. R. (1964). A suggested alternative formulation in the development of Hursch, Hammond, and Hursch and by Hammond. Hursch and Todd. Psychological Review, 71. 528-530.

U.S. Army Training and Doctrine Command (1989). Army training 2007 (TRADOC PAM 350-4, Coordinating Draft). Fort Monroe, VA: Author.

U.S. Army Training and Doctrine Command (1991). Blueprint of the battlefield (TRADOC PAM 11-9). Fort Monroe, VA: Author.

Wallsten, T. S., Budescu., D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1988). Measuring the vague meanings of probability terms (Research Note 88-56). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A196 944)

Wallsten, T. S., & Sapp, M. M. Strong ordinal properties of an additive model for sequential processing of probabilistic information. Psychological Bulletin, 41, 225-253.

Wickens, C. D. (1984). Engineering psychology and human performance. Boston: Scott Foreman and Company.

Youssef, Z. I., & Peterson, C. R. (1973). Intuitive cascaded inferences. Organizational Behavior and Human Performance, 10, 349-358.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems-Man and Cybernetics, 3, 28-44.

Appendix A

Theoretical Framework for Military Judgment

Most military battlefield decision making is a case of cascaded inference, or a dependent series of judgments (see Youssef & Peterson, 1973, for an introduction to cascaded inference models). For example, consider the task of a battalion commander conducting a defense-in-sector. He is in the position to observe the battle on the main avenue of approach (AA), according to original intelligence estimates. However, there are other avenues of approach (AAs) into his sector as well. The commander makes a series of inferences which lead to tactical actions, the first of which is based on the uncertain information gathered from various intelligence assets. His first response to these intelligence data may be to arrange his defenses in a manner that obstructs all AAs into his sector in depth. Simultaneously, he is receiving communications over the radio network from scouts, artillery observers, and commanders which contain estimates on enemy strength, position, actions, and losses. These estimates yet serve as input to other judgments he makes on how best to adjust his defense. Preparing the defense will likely be based on an inference as to the most probable avenue of approach the enemy has chosen for its main assault, and if, when, and/or where a second enemy echelon could appear. These inferences are likely to drive judgments concerning many tactical parameters, such as how the commander commits his reserve forces, or priorities for use of artillery.

There are many models of expert judgment that can serve as a framework in which to illustrate the relationship between a military expert and complex military phenomenon. Most of the judgment models have been developed and evaluated within the context of multiple linear regression, normative theory, functional measurement, and conjoint measurement (see Budescu & Wallsten, 1979; Dawes & Corrigan, 1974; Slovic, Fischhoff & Lichtenstein, 1977; Wallsten & Sapp, 1977 for reviews of these topics). The research paradigms employing these models have focused primarily on choice and inference situations. The choice situation is characterized as one in which a decision maker is presented with two or more stimulus dimensions, and must choose on the basis of the values on these dimensions one of several alternatives for some purpose. The inference paradigm typically presents sampled stimuli and the decision maker must either: (a) decide which of several possible alternatives is true, given the sampled information, or (b) generate a point estimate.

The alternative possibilities in these models are typically represented as being mutually exclusive events. Formal theories of judgment and decision making presume that: (a) the judge has a clear and total picture of the states of the world, (b) the judge also has a clear and total picture of the actions/alternatives that are available, and (c) the judge understands the costs and payoffs for selecting a particular alternative over that of another. With an assumed complete and total knowledge of the world, the judge selects the judgment alternative that maximizes the judge's utilities, or the subjective worth of the judgment (Edwards, 1977).

However, real world judgment tends not to be easily characterized as such a simple single stage process, but rather are multistage. Single stage inference models often lack the complexity for capturing the richness and intricacies present in natural decision environments (Dawes & Corrigan, 1974). For example, in single-stage Baysian inference, probabilities are essentially viewed as prior estimates that are revised as additional information is brought to bear on the decision problem. Here, alternative judgments about the state of the world are based on the Baysian step of revising probability estimates, and when forced to choose, selecting the most probable of all possible judgment alternatives.

Edwards (1977) points out that while single-stage inference models have been extensively studied over the years, they are limited in their ability to capture the essence of decision making in real settings. Furthermore, the reason for this is that the assumptions necessary for applying these models can not usually be met in real world events. Military decision environments profoundly complicate single-stage Baysian modeling because of the following problems (from Edwards, Lindman & Savage, 1963):

(a) In real world cases, data and hypotheses cannot typically be precisely defined and specified.

(b) The judge or decision maker does not usually have the capacity to assign numeric probabilities to the various judgment hypotheses about the world. Further, a judge typically does not follow the rules of probability theory when manipulating and assigning event probabilities.

(c) Probabilities are not stationary and thus the assumption of conditional independence, which is an assumption for using Baysian methods, does not hold.

Military commanders are usually faced with making judgments as to likelihood of complex hypotheses as opposed to the simple hypotheses that typically characterize laboratory-based decision making studies. Rendering judgments about the likelihood of complex hypotheses or scenarios (from a military viewpoint) is complicated, in part, because the decision problem is temporally bound. Possible scenarios evolve and change over time, thus, making it much more difficult to link observed data to the population of possible scenarios. In addition to the temporal characteristics of the decision environment, further complexities emerge from the uncertainty associated with observable data themselves.

Modeling Judgment Output

One descriptive and normative approach to viewing the expert's job at making complex interpretations based on observable data has been elaborated in Brunswik's (1952) Lens model. Brunswik's lens model gave recognition to the importance of natural variability in the environment as a source of variability in behavior. Recognizing the probabilistic nature in natural decision making environments allowed Brunswik to configure a model that would, in part, address some of the concerns raised above by Edwards of precisely defining the structure of the decision problem.

The lens model originally emerged as a means for scientifically representing a complex phenomenon without the need for many of the controls, which Brunswik believed were artificial and superficial, on the environmental conditions under which behavior is observed. Although the model was proposed by Brunswik as a complete model of behavior, it was later modified and restricted for use in judgment processes (Hammond, 1966, 1975). The restricted form of the lens model, however, has maintained the language originally used by Brunswik in his studies of human perception. Research on perception served as the edifice wherein the features of the lens model evolved.

The restricted lens model has traditionally been presented as distinguishing and characterizing the relationship between a judgment criterion that is defined by various stimuli (cues), and the psychological representation of the criterion which is defined through a particular judgment policy. In this case, the concept of judgment criterion is the analog to what Brunswik meant by the term "environment", in the perceptual sense. In extending this notion of the environment in the restricted model, a decision maker produces a judgment of a criterion variable, which is a linear function of a set of information cues. The judgment of the criterion variable that is rendered by the decision maker is based on the judge's personal policy for weighting and integrating information cues in an manner thought by the judge to maximumly predict the criterion variable. The lens model portrays the criterion variable as a function of a series of cues whose relationships with the criterion are less than perfect. A decision maker is viewed as interacting with the criterion, which represents the true state of the world, through a "lens" which is distorted because of the imperfect relationship between information cues and the criterion variable. The relationship between the cues and the criterion variable is typically characterized by "ecological validates" (i.e., zero order correlations) that, in theory, can range in absolute value from 0 to 1.0. Ecological validity represents the predictive importance of each cue.

Figure A-1 illustrates the restricted structural form of the lens model. Here the model defines how a judge uses cue information in making predictions of some criterion variable.

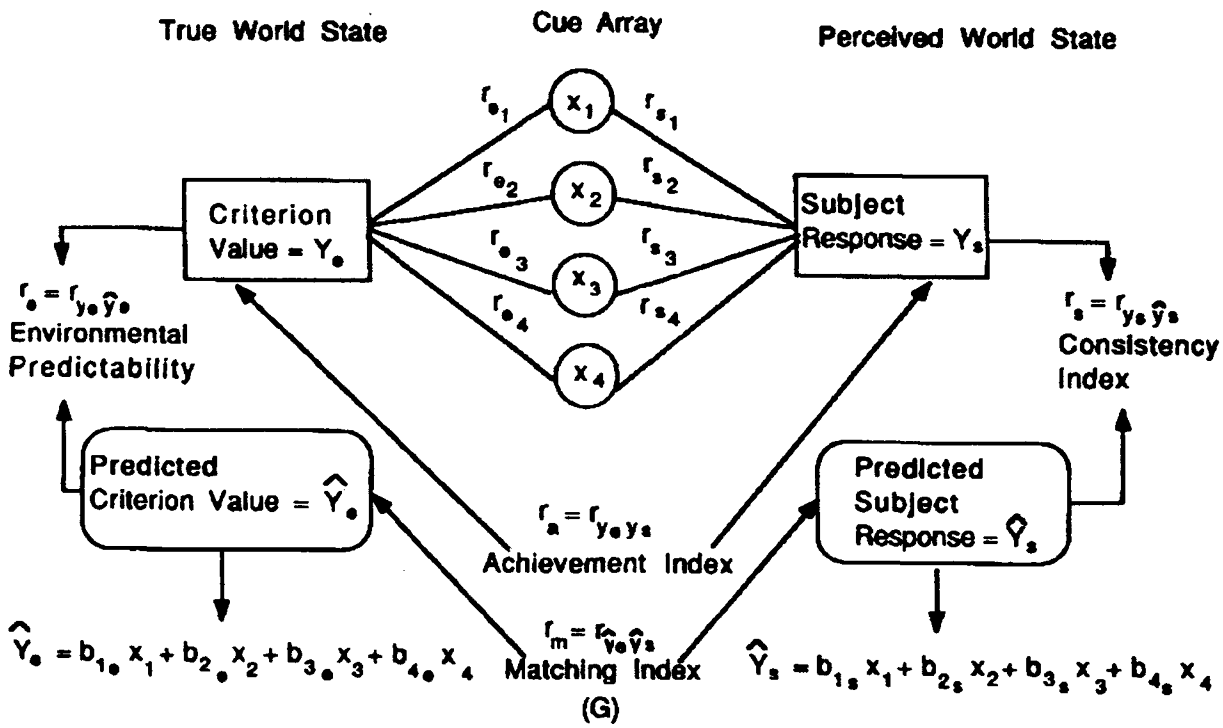

Figure A-1. Restricted form of the lens model.

There are two sides to the model. One defines the true world state, which depicts the truthful relationships between cues and a criterion variable. The other side of the model depicts the perceived world state, which is characterized in a judge's policy for weighting and integrating the cue information. Several indices can be computed which measure the extent to which the judge has a factual representation of the true state of the world. If, for example, the judge produces judgments of the criterion that perfectly correlate with the true criterion values, the judge is said to have perfect achievement. That is, variation in the criterion is perfectly captured by the manner in which the judge utilizes the cue information. However, achievement can be moderated by both a judge's consistency at making judgments (i.e., consistency index) and the judge's representation of how the cues are rank ordered (i.e., matching index). For example, the judge can be very consistent at making judgments of the criterion, but be consistently wrong because the perceived rank ordering of the cues inaccurately represents the true world state. On the other hand, a judge can have perfect knowledge concerning the rank ordering of cues, but, yet be unable to consistently integrate the information in producing criterion judgments.

The lens model has also been applied to the mediation process in policy conflict resolution (Hammond, 1973). The cognitive conflict paradigm is defined as a situation in which two or more parties are trying to solve a common problem, and conflict is caused by differences in judgment policies. Here, the criterion side of the model is replaced by another judge. The essential notion pursued in the conflict mediation paradigm is that two or more judges viewing the same information may display differences on how that information is used in developing a judgment policy on some issue. The model allows one to document discrepancies between judges in an effort to mediate and ultimately resolve policy disputes (see Hammond, Stewart, Brehmer & Steinman, 1975).

Building on the idea of conflict resolution, the lens model appears to be a suitable candidate for representing the structural similarities and differences in decision making characteristics between a military performer and an expert military observer. The expert observer is considered to possess a doctrinal model of the tactical decision problem which is embedded within a network of military science propositions. The performer also has a model of the tactical situation and a knowledge-base which serves as a means for interpreting field data. The lens model provides a theoretical framework and methodology for examining, in detail, a military performer's policy for selecting and integrating tactical information thought by the performer to optimally predict various tactical outcomes. The perfomer's policy can be contrasted, in theory, with the judgment policy of a military expert in an effort to help train performers on the efficient use of tactical information.

Within the context of the lens model, the performer faces several challenges in processing a multitude of data sources that link with alternative tactical actions. For example, each data source is not perfectly correlated with the tactical criterion, such as achieving a successful movement-to-contact mission. Instead, each data source differentially predicts a component of the criterion. This is to say that the tactical criterion is multidimensional in nature. Among the tasks the performer must accomplish in producing statistically optimal judgments are (a) selecting those data sources that are important in predicting the outcome of a tactical operation (i.e., choosing relevant information); (b) applying expertise in determining the extent to which each data source is predictive of a given tactical outcome (i.e., deciding how relevant each information source is); and (c) integrating all the information on the basis of its diagnostic value for predicting a tactical outcome in order to make a judgment as to the likelihood of that outcome (i.e., deciding how to combine all the information to yield a judgment).

The model has both descriptive and normative features which help to evaluate the performer. First, the model allows one to describe the information sources that were selected by the performer to be most predictive of a particular tactical outcome. Secondly, the importance (i.e., the weight) assigned to each information source by the performer can be assessed. Finally, the manner in which the performer integrates the information can be described. For example, is the performer's judgments most predictable from a predominately additive linear model, or is a configural nonlinear model better able to capture the manner in which the performer used the information?

The normative properties of the model allow for contrasting the judgment dynamics of the performer with those of the expert observer. The expert observer's judgment model of the tactical situation serves as a standard in which to evaluate the judgment skills of the performer. Several statistical indices (mentioned above) document different dimensions of the comparison between performer and observer.

For example, the manner in which a performer uses particular cues can be modeled by a regression equation that predicts the performer's judgment of the expert observer's evaluation from a linear combination of cue weights. The degree to which a performer accurately assesses the characteristics of the expert judge is expressed by the correlation between the expert's judgments of a tactical criterion and those judgments predicted by the performer.

The lens model can be mathematically characterized by defining the relationship among the components of the model for the expert judge and judgment task performance on the part of the performer. Tucker (1964) described it as follows:

$$r_a = GR_sR_e + C[(1 - R_s^2)(1 - R_e^2)]^{.5}$$

The correlational performance an individual achieves (i.e., achievement index) r_a, is a function of four distinct components. Component 1 is the linear multiple correlation between the cue values and the expert judgments, R_e. In the original language of the model this would be termed environmental predictability. However, following the policy mediation concept, environmental predictability is replaced by the notion of predictability of the expert judge. The index essentially characterizes the uppermost predictability of the judgment task. Component 2 is the linear multiple correlation between the cue values and a performer's judgments of the expert's evaluations, R_s, (consistency index), which represents the ability of the performer to control the execution of the judgment policy he believes is also being used by the expert. Component 3 is the extent to which the linear model of the performer's judgments correlates with the linear model of the expert' judgments, G, (matching index), which measures the performer's task knowledge. Finally, Component 4 is the extent to which the nonlinear residual variance in the model of the performer correlates with the nonlinear residual variance in the model of the expert, designated C.

The lens model has several limitations that have been extensively discussed in the literature (see Dawes & Corrigan, 1974; Slovic & Lichtensein, 1971). Two of the more prominent complaints have been associated with the model's limited robustness and failure to address the intervening processes that lead to judgment.

While the multiple linear regression model has been relatively successful in reproducing a decision maker's judgment policy, it may not reflect the underlying cognitive processes of judgment. The regression model may be successful at fitting a decision maker's behavior because of its robustness in face of nonlinear relations and variations in the cue coefficients (Dawes & Corrigan, 1974). Further, the robustness of the model makes it difficult to disprove.

Secondly, the lens model is essentially an output model. It emphasizes the input-output characteristics of decision making, but is limited in making inferences about the intervening processes of judgment. That is, it is relatively insensitive as a method of discovering, testing and explaining what goes on between the presentation of the cues and the performance of the responses. Judgments are evaluated for efficiency and optimality on the basis of statistical criteria. Thus, the focus is on the final judgment, not how the judgment was formed. However, making the claim that a judge acts like a particular statistical algorithm is clearly inappropriate. What we are free to say using the lens model is that a certain model is best for statistically capturing the judgment behavior of the performer.

Modeling the Military Judgment Process

Another conceptual framework for how an expert military observer processes tactical information in selecting among potential judgments can be presented as a state analysis problem. A process model of judgment provides a framework for examining the details associated with the rules and mechanisms that are the antecedents to judgment. The process model is a representation of the judgment policy itself, where elemental features of the policy are exposed.

One conception of a state system is based upon that adopted in the theory of dynamic systems and intelligent automata (Arbib, 1972). The state concept is used here to discuss the judgment process as a sub-set of biological systems (Bunge, 1980). The emphasis is on knowledge states or cognitive awareness states, rather than a state defined by the outcome of some global inference, such as that obtained with the lens model formulation.

A complex situational awareness system may be viewed as existing at any given point in time in one of a large number of states. From a theoretical viewpoint, the number of states can be infinite. The state space is defined as the set of all states the system can be in, and is represented by an n-dimensional array made up of the functional ranges for each property of the system. A particular state is defined as a point in this space which is represented by a pattern of values that correspond loosely to what is sometimes called the "estimate of the situation". In this case, the properties that define situational awareness can be considered system vectors. From a practical standpoint in modeling situational awareness, the number of properties, or indicator variables, is kept relatively low. Only variables thought to be important in determining the system's behavior are considered.

An additional feature that is important to consider is that in any dynamic system the state space will be constant flux. It is probably unlikely to be either valuable or possible to consider transient states that endure only briefly. Furthermore, one may argue that as expertise develops, the ability to quickly categorize information becomes better. One might expect that this would lead to system stability and reductions in fluctuation of the system.

Figure 4 presents a finite state model of tactical judgment. There are essentially four basic components shown in the model: (a) the tactical environment, (b) the human observer/performer, (c) the state of situational awareness, and (d) the action space. The integral idea is one recognizing and processing meaningful patterns of data in the tactical environment, and mapping the patterns of meaningful data over to the action space for the appropriate judgment. The process is entirely mediated by situational awareness. This conceptualization of the process was suggested by Bessemer (1991b).

In the state process model, the tactical environment can be partitioned into two subcomponents. The first subcomponent may be envisaged to contain apriori information that remains relatively static during the battle. This data category can describe an almost unlimited amount of information as long as it is historical in nature. For example, it can include past battlefield intelligence, historical knowlege of enemy military doctrine, formulated battleplans, orders, fragmentary orders, and standard operating procedures. This apriori knowledge is the supporting context for real-time decision making.

The second subcomponent of the tactical environment symbolizes real-time data elements that are evolving around the observer. These real-time activities occurring on the battlefield and in communications, in theory, condition or modify dynamically what is known about the current situation. For example, the validity of past intelligence information is strengthened or weakened on the basis of the information now available to the observer. The combination of static apriori and changing real-time data represents all the potential information that the judge can draw upon in making judgment assessments of the tactical situation.

Clearly, there are certain physical attributes that must be present in the human component of the model for acceptable judgment performance. Although we assume the observer's or performer's physical senses are intact and performing optimally, this is certainly a simplifying assumption. A more comprehensive model would include a provision for state changes associated with many physiological system parameters as well. In fact, it can

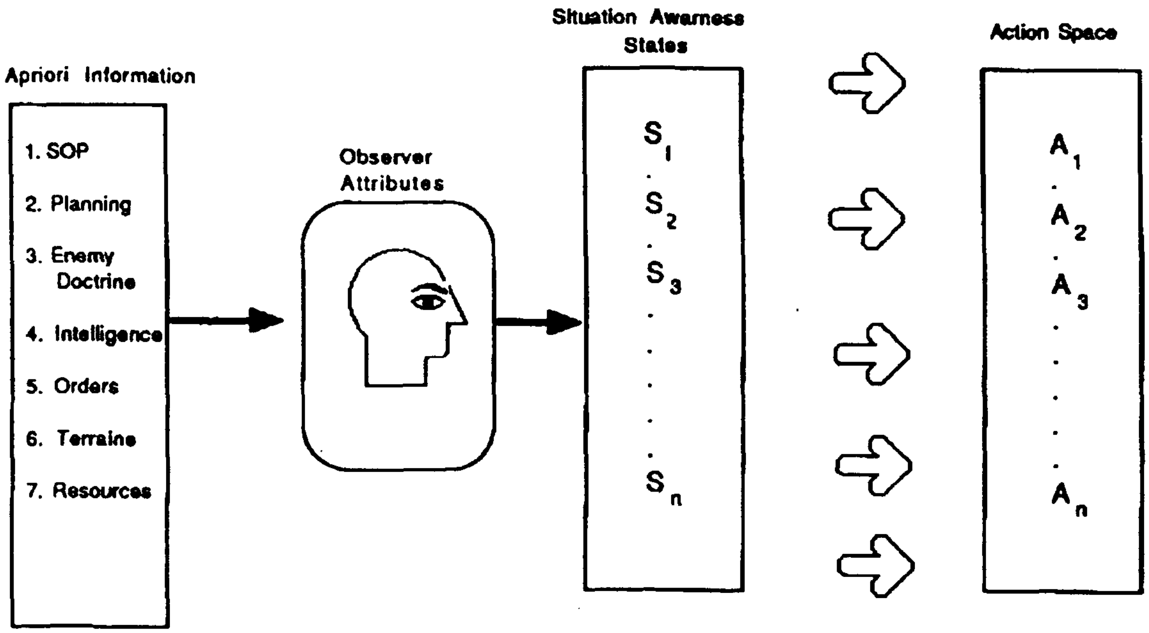

Figure A-2. State judgment model showing four components: tactical environment, human observer/performer, awareness vector, and action space.

easily be argued that situational awareness is conditional on the many fluctuations in the physiological state of the observer.

However, in the interest of limiting the scope of the discussion, we focus on a situational awareness system that can be defined, at least in part, through military science propositions. The basic concept of situational awareness fits rather well in a descriptive framework approach to modeling complex tactical judgment. However, there is some inherent ambiguity in the term that gives rise to multiple meanings and thus uncertainty about how it is to be conceptually defined, and how one goes about measuring it. For the present, we restrict the discussion by loosely defining the concept along the lines considered by military commanders when speaking of intelligence preparation for the battlefield. However, in this case we also consider real-time interactions with the battlefield.

In the context of decision making within a temporally bound, rapidly evloving environment, certain decision behaviors undergo modifications as new information becomes available. There are a certain number of assumptions that come with any military operation, and this helps set the stage for guiding a decision maker's actions. Within this state framework conception, these fundamental assumptions are associated with (a) what is known about the tactical environment now and (b) what is known about military science constructs and propositions. These elements will tend to influence the decison making process. However, the tactical environment remains in flux to some extent, so the decision maker must always be updating what is known with respect to the events unfolding before him, and how this information will influence the application of certain military ideas.

It is probable that situational state vectors, in part, would be "tuned" to historical and doctrinal parameters, such as weather and climate, terrain, enemy forces, troop availability, and movement routes, that are typically considered during battlefield preparation. Much of this knowledge is captured in the notion of METT-T factors (mission, enemy, terrain, troops availability and time). Here, the situational awareness component of the state model can be conceptualized in terms of constellations of indicator variables which occur together in well defined patterns. While it is clear that unplanned events external to events considered during the production of orders and fragmentary orders will occur, it is likely that these historical events will set the thresholds for the awareness state, which itself responds moment to moment during battle.

Consider the example of a platoon leader engaged in a military operation. He has access to various apriori historical information in the way of orders, intelligence, unit strengths, resources, plans, and so on. This knowledge, in part, conditions the awareness state by preparing the leader to look for specific events during the battle. As the battle evolves, the pattern of values on the dimensions making up the awareness state change as new information becomes available concerning the tactical environment. These moment to moment changes follow fluctuations in the leader's attention to certain features of the battle. In this conception of situational awareness, each state is described by a unique vector of values in the multidimensional awareness space. Here, the vector combines the values for each of the system variables that the platoon leader momentarily determines to be possible influences on the outcome of the battle.

The final component of the model characterizes the action space, or the judgment alternatives available to the platoon leader. A knowledge-based rule would link a particular awareness state configuration to the action judged best, given the situation defined by the information being attended to by the platoon leader. For example, a particular array of values of the vectors making up the awareness state may lead to the action; "send situation report now". However, another array of values, which presumably reflect a different tactical situation, would lead to a different action, such as sending a movement command to the platoon. At any particular time of the battle, the conditions necessary for several mutually exclusive actions may be possible. These actions form the action space for the model. While only one action is possible at any give time, a particular state space configuration can establish the necessary conditions for more than one action. That is, a given awarness configuration can map to more than one action. However, in theory these other actions will have to be deferred until that action which is judged to have the optimal outcome is completed.

Appendix B

Future Research

This report has discussed some of the theoretical issues associated with military judgment, and the need for a more flexible measurement theory that can link objective and subjective indicants of performance to the constructs of military science. We are now in the position to outline a research approach and plan that illustrates one possible path leading from the theoretical implications of fuzzy sets to practical applications of fuzzy set methods for performance measurement in simulated military training systems.

The goal in developing a useful performance measurement system would be to create a set of procedures whereby both numeric and linguistic parameters of a military exercise could be measured directly or otherwise estimated. The measures would then be transformed and manipulated according to rules that provide a valid means for combining this information in indices representing the major dimensions of performance. It seems appropriate to consider the potential role of a UPAS-like measurement system for producing various quantities that represent degrees of the qualities of important military dimensions, such as those presented in the results section of this report.

Several techniques for producing this information would be possible. For example, UPAS could be configured in such a way as to record various objective indicants that would serve as input to a series of algorithms that connect the objective information to linguistic performance variables. The linguistic assessments could be used in a variety ways by instructors for feedback in after actions reviews, and as a means for transmitting training quality control information to managers. The significance of this approach would be to retain the descriptive linguistic metrics. In theory, such metrics could summarize and communicate complex information in a manner more compatible with the way military observers understand a given military problem. However, for some diagnostic purposes, it may be appropriate for UPAS to present actual numeric scale values which would represent the degrees of qualities for militarily meaningful dimensions. In this case, an algorithm translating linguistic values into numeric values would be useful to display the precise meaning conveyed by the graded and more vague verbal descriptions of the military dimensions.

In order to make it possible to use and manipulate linguistic information, rules need to be generated that map the objective indicants of performance over to linguistic values that represent a consensus of expert judges' verbal descriptions of various military dimensions of interest. This step in the process of applying fuzzy techniques is the most difficult, and is essentially the core feature of any program of research investigating a measurement system of this nature. Further, it clearly incorporates at least two stages: (a) discovering and developing a semantic network that is representative of the natural language system used by expert military judges in describing various dimensions of performance, and which is the focus of the case study in the present report; and (b) estimating the quantitative membership functions that characterize the grade of membership of the elements of various physical parameters into the linguistic categories, and that make possible the use of fuzzy operations on these categories.

While there appears to be many approaches available that would seem to adequately document the semantic networks used by experts in verbally describing military performance, such as the method of "progressive elaboration" illustrated in this report, the issue of membership functions is much more difficult to address. Defining membership functions for linguistic variables can be approached in a number of ways. In some applications, the functions are essentially arbitrary in the sense that they are tailored to be useful in a specific domain. For example, Schumcker (1984) indicates that in the area of risk analysis, the computer system designer often produces functions that are intuitively meaningful, and that the designer believes will serve to adequately communicate the normal meanings given to the English terms represented in the syntax. In other applications, it becomes more appropriate to empirically derive these functions in order to reduce errors associated with arbitrarily selecting particular membership values for a given function.

Smithson (1987) details the pros and cons of various approaches to empirically deriving membership functions. Essentially, Smithson (1987) indicates that a universally accepted methodology does not yet exist. Each approach has advantages and disadvantages. However, the important point to consider for the purpose of research and development of a measurement system is that a fuzzy methodology for producing membership functions from judgment and rating data can be validated empirically. Demonstrating that relations among various membership functions generated from different kinds of judgments are both reliable and valid permits direct tests to be made on the axioms, definitions, and theorems that represent the foundation of fuzzy set theory.

Since there are many claims that exist as to the specific manner in which fuzzy sets should be transformed (see Smithson, 1987; Zadeh, 1973), it seems the first logical step would be to examine some militarily relevant fuzzy sets. A candidate system of inquiry might be directed at determining fuzzy properties of communication performance outlined in the case study section of this report. Generating baseline information on the fuzzy sets associated with the communication dimensions discussed above would permit evaluating the transformations that occur when these sets are operated on by various set operations, such as negation, conjunction and disjunction.

Several potential experiments seem to be appropriate at this stage. The goal would be to document how predicted fuzzy operations would correspond with the operations actually performed by military judges. For example, a series of physical measures might be generated that would represent the objective data for assessing the qualities of the communication constructs by experts. As a concrete example consider the "timeliness" construct above. A set of message times could be evaluated in terms of their degree of membership in several linguistic categories, such as the ones associated with the timeliness construct in the results section of this report.

The actual process of determining membership could be made very simple, or complex. A simple approach would be for the experts to indicate whether a message time value was a member of a particular group of fuzzy sets. This approach has been used in the past by Labov (1973) and others. The grade of membership of a particular message time value in a given fuzzy set would be a function of the proportion of experts indicating that it indeed belonged to the set in question. This approach to deriving membership values has several critical drawbacks that Smithson (1987) argues limits its validity. Another approach that Smithson (1987, p. 81) presents as having advantages over the previous method is having experts make rating or ranking judgments which can be analyzed with conjoint, multi-dimensional, or other scaling algorithms. The drawback to this approach lies in its complexity associated with computing the inter-stimulus distances from the judgments in order to establish membership in fuzzy sets (see Smithson, 1987).

However, once membership functions defining the fuzzy sets have been recorded, the functions can then be compared with predicted functions resulting from operations specified in fuzzy set theory. Zadeh (1973), Smithson (1987), and others give specific arguments in favor and against particular functions for given fuzzy sets. While there exists a certain amount of agreement on acceptable functions and operations for some linguistic hedges and other modifiers, there is still debate over the transformational properties of many fuzzy sets. However, Zadeh (1973) and Smithson (1987) provide functions for some of the more general fuzzy sets.

A second type of experiment could be configured to address a feature of the standard approach used in developing membership functions (see Labov, 1973). Several methods for measuring membership in fuzzy sets essentially present a judge with all of the linguistic categories and then ask the judge to make confidence ratings, yes/no judgments, subjective assessments and so on of the membership characteristics between the stimuli and the fuzzy sets (Labov, 1973; Zadeh, 1973). However, it is possible that this creates a confounding context effect similar to that effect observed in psychophysical scaling experiments which may alter the true nature of the membership functions under study. For example, the impression of differing degrees of weight of objects may be affected by the average weight of objects that make up the group to be scaled. Thus, scaling very heavy objects may produce relative psychophysical functions that are very different than the functions generated from a group of very light objects. Similarly, the membership functions generated from a set of message time values may be anchored in some way by the relative values in the fuzzy sets selected by the experimenter. Therefore, experiments that show how to remove, control, or measure this type of contextual effect may help to demonstrate the robustness of a set of membership functions for a given set of linguistic categories.

Rather than present all of the linguistic categories to each judge, we are interested in presenting only one category to each judge. For example, each judge would be presented with only one category or fuzzy set and asked to pair, by which ever membership derivation method selected, those values applicable to the fuzzy set in question. This would help determine whether or not a scaling bias associated with conditioning responses on the set of possible responses in the experiment existed. The membership functions derived would be compared with those functions predicted from the axioms of fuzzy set theory.

The outcome of a set of experiments documenting the correspondence between the operations specified by fuzzy theory and those obtained from an empirical investigation of expert judges evaluating dimensions of military performance would lead to domain dependent hypotheses about set relations. That is, once the question about valid membership functions and operations is answered, one is then in the position to examine the effects of other independent variables on expert judgments. Appendix C illustrates a potential set of hypotheses concerning the fuzzy definitions of the communication constructs presented in the case study section of this report.

Clearly, one major objective for a program of research on fuzzy performance measurement is to develop software to collect and process judgment data complementary to the objective measures obtained by a UPAS-like system. This will mean determining, among other things, the manner in which judgment data is collected from expert military observers. Similarly, the fuzzy algorithms that process this information will have to be deveoped simultaneously since the input information must be formatted in a manner compatible with the algorithms. Furthermore, the display formats for presenting fuzzy information for various training purposes will become important to consider, and how that data can be usfully manipulated for further examination by the end user.

Appendix C

Relations Among Concepts Describing Reports

The terms that are used to describe military communication performance are related to three aspects of the communication process: (a) situation, (b) time, and (c) message content. By examining the reporting process as one special case of military communication, several hypotheses will be developed about the relations among terms that describe each aspect. The hypotheses begin to sketch out the framework for a fuzzy theory of reporting performance. The main objective for developing such a theory is to establish a foundation for the measurement and evaluation of the sender's performance. Hopefully, such a theory could then inspire analogous theories for other forms of communication. However, the specific hypotheses concerning reporting may or may not generalize to other types of communications.

Three bipolar dimensions are suggested as the essential properties that distinguish variations in the sender's performance among individual reports. First, reports may be necessary or unnecessary depending on the situation. Second, reports may be timely or untimely depending on when the report is transmitted. Third, reports may be informative or uninformative depending on their content.

A basic working hypothesis adopted here is that all evaluative terms that refer to reporting performance are interpretable as some direct function of these three dimensions. For example, "good" reports are necessary, timely, and informative. If the positive term of the bipolar dimensions are measurable fuzzy sets (N, T, I), the hypothesis about the set of "good" reports (G) should be testable using the various formulations of fuzzy intersection (min/max, product, or bounded sum). In this case, the negative terms for each dimension are construed as the fuzzy complements (N', T', I') of the positive terms. Using the usual symbols for set intersection ($\cap$) and union ($\cup$), the hypothesis for evaluation on a good-bad dimension is: $G = N \cap T \cap I$. Assuming that "bad" is the complementary fuzzy set $B = G'$, then $B = N' \cup T' \cup I'$.

The next question is how membership functions for the three fuzzy categories and their complements should be measured. A direct approach through some membership scaling method might be attempted to address this question. However, further consideration of what an observer of the communication process might take into account when asked to judge reports suggests that these dimensions are complex functions of other more elemental properties of reports. Each bipolar dimension relates to (at least) two other properties as listed below:

1. Necessity
   a. Possibility
   b. Priority
2. Timeliness
   a. Promptness
   b. Brevity
3. Informativeness
   a. Completeness
   b. Accuracy

In the sections that follow, the communication process is examined in detail to elucidate the relations among these concepts. For simplicity, the concepts and their relations will be developed considering only the sender's involvement in the process, independent of the receiver's performance and the interaction between sender and receiver. Since the object of the present exercise is to develop a means of individual evaluation, we are driven to limit the scope of the inquiry to the sender's contribution to the process. From this standpoint, the receiver simply becomes another part of the environment to which the sender must react.

Situation

At any point in time, a military leader must be prepared to anticipate and react to a multiplicity of factors that influence decisions in a tactical situation. As an information processor, the leader can be loosely conceptualized as operating as a finite state machine. His current state of situational awareness can be represented as a state vector of variables $(v_1, v_2, \ldots, v_n)$ that include elements of information that are either currently in the focus of attention or directly available in immediate memory. Some of these elements are perceptions or concepts associated with factors in the current environment. Other elements are retained from preexisting factors, such as plans, SOPs, and prior events. As time passes, the leader actively searches his environment for additional information, monitoring information sources for changes. As changes take place and are detected and recognized (i.e., events occur), the leader's state of awareness for situational factors changes. Changes in the current values of variables in the state vector represent changes in awareness.

Every action, including reporting actions, that the leader might perform has a number of necessary conditions (initiating conditions) that should be satisfied to make the action either possible or desirable. If an action is performed lacking the necessary conditions, this action is regarded as an error of commission. If action is not performed when the necessary conditions have been satisfied, this is regarded as an error of omission, unless other circumstances exist that make the omission appropriate. The links between conditions and actions can be represented as relations defined between patterns of values in the state vector and the set of alternative actions. The assumption is that the initiating conditions for all actions are included among the values of variables in the state vector.

At any particular time, the initiating conditions may be satisfied for several mutually exclusive actions that cannot be performed at the same time. Such actions form a set of possible actions that can be performed, but only one action can be started, while the rest remain pending. Thus a full production system modeling the leader's choices among actions must include mechanisms that order the priorities among pending actions, and only the action that gains first priority is actually performed.

In general, the initiating conditions for reports are some pattern of values in the state vector that include the elements of information to be reported. Other conditions may also be required to make the report possible. For example, a phase line may be named DOG in the mission order and a unit route indicated on the map overlay. When the leader is aware of his location, then the information for a location report is available. If the SOP dictates a location report be sent when the unit reaches a phase line, then proximity to the phase line is also required to make the report ("DOG, now") possible. However, the report is not sent at once if other actions are also pending with a higher priority, e.g., issuing an order to change the unit formation.

In this conceptualization of the communication process, a report becomes possible at time t_0, when the sender acquires the information, and other initiating conditions for the report are satisfied. This time is always somewhat after the objective time, t_I, when the initiating conditions were first satisfied, since detecting and recognizing the information requires some time. The possible report then remains pending until time t_1, when the report gains first priority among all other pending actions. Within this framework, a report is necessary if and only if it is a possible action with first priority. Neither t_0 nor t_1 are objectively measurable, and cannot be determined with precision by an observer. Therefore, both times are fuzzy variables from the observers' standpoint. Only the start and completion of the report transmission are events with objectively measurable times t_R and t_C. An observer's judgment must be based solely on the latter times and his own situational awareness of factors that affect the possibility and priority of pending actions. The hypothesis is that a report will be judged necessary when it is judged to be possible, and when all other pending actions judged to have higher priorities are completed.

If C_X is the fuzzy set representing the observer's judgment that action X is possible, i.e., that conditions for action X have been satisfied, and F_X is his judgment that action X has highest (first) priority at the time it is performed, the necessity hypothesis requires that the observer judge the report (R) both in relation to the series of actions performed before the report (B[1], B[2],..., B[n]) and in relation to the actions pending at the time of the report and usually performed afterward (A[1], A[2],..., A[m]). Then the hypothesis is as follows:

$$N = \bigcap_{i=1}^{n} (C_{B[i]} \cap F_{B[i]}) \cap (C_R \cap F_R) \bigcap_{j=1}^{m} (C_{A[j]} \cap F'_{A[j]})$$

This complexity of the hypothesis is dictated by the complex dependency of any sequence of actions on multiple aspects of an existing tactical situation. Rather than being overly complex, the hypothesis undoubtedly oversimplifies the difficult problem of judging when a report becomes necessary within an ongoing rapid sequence of actions.

Time

Current tactical doctrine suggests that reports should be transmitted as rapidly as possible when they become necessary. This implies that the transmission should be initiated promptly (with minimum latency) and should be completed as briefly (with minimum duration) as possible. In theory, timeliness should be inversely related to the total time (t_T) required to complete the report, or $t_T = t_C - t_1$. If promptness is inversely related to latency, $t_L = t_R - t_1$, and brevity is inversely related to duration, $t_D = t_C - t_R$, the fact that $t_T = t_L + t_D$ implies that some relatively simple and systematic relation should be found between the fuzzy sets representing the concepts of timeliness (T), promptness (P), and brevity (B). The obvious hypothesis is that $T = P \cap B$, with either the max/min, product, or bounded sum form of fuzzy intersection. On the other hand, if t_L and t_D are weighted unequally in judging timeliness, some more complicated relation may be found.

One difference between promptness and brevity is that the observer must infer one of the values (t_1) that affect t_L, while both values determining t_D are directly observable. Much of the information about the situation available to the sender of the report is not usually available to the observer, so the latter's estimate of t_1 can be expected to be error prone. Typically, the error will tend toward underestimation, since the observer often will be unaware of one or more of the sender's priority actions that delay reporting, and thus increase t_L. This should make promptness fuzzier than brevity, tending to reduce the membership values at longer times. Therefore the fuzziness of timeliness judgments may be found to relate more strongly to the promptness membership function rather than that for brevity. If the intersection hypothesis is correct, the membership values for promptness will tend to dominate timeliness with either the min/max or product form of intersection.

In training exercises conducted in field or simulator settings, the observer often will know that the reporting conditions were satisfied at t_I. He may also know the location and approximate orientation of the report sender, and can form an estimate of t_0, by which time the sender ought to have acquired the information for the report. If the observer can monitor communications within the sender's vehicle and on the unit net, he has a partial basis for inferring what the sender is doing, what some of the pending actions are, and the priorities among them. These indications help the observer to form his estimate of t_1, when the observer thinks the report should have become necessary. However, under the best of circumstances there can only be a loose relation between the observer's estimates and the actual behavior of the sender.

Other message traffic on the command net introduces an additional complication. If the sender's access to the net is blocked by conflicting transmissions, t_R will be increased, thereby reducing the sender's promptness. To the extent that the observer recognizes such incidents and can estimate the added delay (t_X), he should then base his judgment of promptness on the adjusted latency, $t_L - t_X$. To the extent that the observer is not aware of conflicting transmissions, or fails to adjust for their effect, his judgment will be biased against promptness.

Commanders are often found to be impatiently awaiting a report from a subordinate that they consider to be unnecessarily delayed. Since commanders, like any other observer, are unable to fully assess the justifiable reasons for delays in reporting, their judgments of promptness are subject to similar sources of bias. With incomplete knowledge of the situation, commanders will frequently make insufficient allowance for report delays.

Message Content

Some messages, such as spot reports and shell reports, have official names and formal content structures (formats) prescribed by operational doctrine and SOPs. Report formats list specific elements of information (report lines) to be provided in a specific order. Some of the elements in formatted reports are essential defining features for that type of report, while other elements are optional, depending on the available information and the situation. Other reports have no prescribed format (except that required by standard communication procedures) or recognized names, but may be classified by the type of information transmitted. Table 1 shows the named formatted reports that are used most often and other common types of reported information.

A formatted report should be regarded as complete if all of the defining elements are fully transmitted. Usually, a bare minimum of specific information is sufficient to transmit an element. Additional detail is considered unnecessary for completeness, and even undesirable. Overly elaborate detail will detract from the impression of brevity. In extreme cases, an element of information that is very vague may detract from an observer's judgment of completeness. At the extreme, vagueness can be nearly the same as omission of information. However, most of the elements required in formatted reports leave little room for vagueness. On the other hand, the elements of unformatted reports are self-defining, and should usually be regarded as complete if they provide specific information on whatever elements are transmitted. When present, vagueness can be expected to have a greater effect on completeness judgments for these reports.

While completeness judgments can be expected to be based primarily on defining elements, it is also possible that an observer's judgment of completeness could be influenced by optional elements as well. If the observer has some indication that the sender is in possession of the information required for an optional element, then it is likely the report will be regarded as incomplete to some degree if that element is omitted.

Table C-1

Common Types of Reports

Formatted Reports	Unformatted Reports
Contact	Movement
Spot (SPOTREP)	Location
Call for Fire	Navigation
Adjust Fire	Vehicle Identification
Situation (SITREP)	Landmark Identification
Route	Equipment Information
Shell (SHELLREP)	Friendly Unit Information
Ammunition	Enemy Unit Information
Nuclear, Biological, or Chemical (NBC)	Clarification, Update, or Addition to Report

The information elements contained in a report may be values of a categorical variable (e.g., type of vehicle, activity), comparative or ordinal variables (e.g., moving slowly, tanks leading column), frequency counts (e.g., ten tanks), or continuous variables (e.g., time, grid location). Each element of information is regarded as accurate if it corresponds to the actual situation at the time. An observer with full knowledge of the situation can presumably determine the accuracy of each element reported and form an aggregate judgment for the report as a whole. However, deviations from accuracy are measured in different terms for different types of variables, and the importance of the information varies from element to element. For both these reasons, the process of forming an aggregate judgment cannot be very simple.

The role of unreported elements is another source of complication. If it is assumed no accuracy is possible if an element of information is omitted, this corresponds to making completeness a necessary but not sufficient condition for accuracy. In this case, the two concepts are not independent, and a high degree of accuracy will imply a high level of completeness, but the reverse implication does not hold. On the other hand, low completeness will necessarily be associated with low accuracy. Whether observers' judgments will actually obey these relations is presently unknown.

If the observer does not have information available about the reported elements, then a judgment of accuracy cannot be made at the time of the report. He must wait for the situation to develop further, obtaining additional information at a later time that can verify or contradict reported information. Often, obtaining the necessary information for a feature of the situation that is temporary (e.g., a location) may become impossible when its status or value changes after a brief interval.

However the judgments are made, the hypothesis made here is that informative reports are both complete (to some high degree of completeness) and accurate (to some degree of accuracy). If C and A are the fuzzy sets including the required degrees of completeness and accuracy, respectively, the hypothesis is $I = C \cap A$. Among the noninformative reports, some are regarded simply as uninformative (C'), while others are misinformative ($C \cap A'$); thus $I' = C' \cup (C \cap A') = (C' \cup C) \cap (C' \cup A') = C' \cup A'$.

Testing Hypotheses

The concepts and hypotheses that have been developed in the preceding sections lead to a number of research problems and empirically testable questions. The fundamental problem is to estimate membership functions for the polar terms, and to measure consistent membership values for reports in particular tactical situations. Wallsten, Budescu, Rapoport, Zwick, & Forsyth (1988) illustrate one promising method that was applied to probability terms. Their method seems to be readily adaptable to scaling membership functions for other terms.

Given membership values on appropriate scales for a sample of reports, the hypotheses presented here can then be tested directly along with the various definitions of fuzzy union and intersection that have been proposed. Smithson (1987) presents several definitions of fuzzy union and intersection, and discusses previous research related to these definitions.